The Thesis Committee for Yoon Ho Bai
Certifies that this is the approved version of the following thesis:

Human and monkey detection performance in natural images compared with V1 population responses

APPROVED BY
SUPERVISING COMMITTEE:

Supervisor: _____
Wilson Geisler III

Co-supervisor: _____
Eyal Seidemann

Human and monkey detection performance in natural images compared with V1 population responses

by

Yoon Ho Bai

Thesis

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Master of Arts

The University of Texas at Austin

August 2017

Abstract

Human and monkey detection performance in natural images compared with V1 population responses

Yoon Ho Bai

The University of Texas at Austin, 2017

Supervisor: Wilson Geisler, Eyal Seidemann

Detection is a fundamental task that is critical to visual behavior. The central aim of this study was to measure and model behavioral and neurophysiological performance for detecting targets under naturalistic conditions. I first measured behavioral detection performance macaques and compared it to humans. Detection thresholds were measured on uniform backgrounds and for several contrasts of natural image backgrounds. I find that (i) threshold contrast power is a linear function of background contrast power for both humans and macaques, and (ii) the relative threshold functions for humans and macaques are in good agreement, although (iii) the macaques are less sensitive overall. Subsequently, I investigated the quantitative relationship between V1 population responses and detection performance. I used voltage-sensitive dye imaging (VSDI) to measure the neural population activity in V1 for the same stimuli, while the monkeys held fixation. The spatial scale of VSDI measurements was sufficient to resolve retinotopic responses and orientation columns over the whole region activated by the target. Separate read-out strategies were used for retinotopic and columnar responses. Across multiple contrast levels of natural

image backgrounds, I compared both scales of population responses between target-present and target-absent conditions to derive the signal-to-noise ratio (d'), which specifies neurometric functions. Based on this simple approach, the results show that in comparison to behavioral performances, retinotopic performances degraded at a relatively higher rate with increasing contrast masking. On the other hand, columnar performances were relatively less susceptible to contrast masking in natural image backgrounds.

Table of Contents

Chapter 1: Introduction

Human visual perception relies on the successful interpretation of patterns of light that are encountered in the natural world. Light in the physical environment is first transformed by the optics of the eye that focuses light onto the retina. This light is then transformed into electrochemical signals through an array of retinal sensors that establish the initial image encoding that the rest of the nervous system can use to guide behavior. As the retinal image is a projection of a three-dimensional world, it portrays two-dimensional arrangements that are inherently ambiguous. Ambiguities arise because an infinite number of three-dimensional configurations in space collapse to the same two-dimensional image. This presents a challenge to the visual system in selecting features that will help understand what properties of the environment caused the visual stimulation on the retinal image. To solve this challenge the visual system carries various kinds of sensory processing that aid in interpreting the image. Furthermore, the visual system utilizes regularities in the natural environment to make inferences (Knill and Richards, 1996; Geisler et al. 2008; Girshick et al. 2011). The properties of these natural signals have particular statistical regularities that limit how well the visual system can perform. Given these natural constraints on vision it makes sense for the visual system to have evolved to account for these regularities when making inferences about the physical world. Therefore, measurement of these physical properties in the natural environment, understanding how they influence visual perception, and how the nervous system has evolved to handle these natural signals is central to vision science.

This thesis is centered on a scientific rationale suggesting that the design of our visual system is closely related to the statistical properties of the environment (Geisler 2008). Natural scenes are inherently variable with a wide range of structures and spatial

features. Nonetheless, the visual system has adapted to exploit reliable sources of information through statistical inference to extract properties of interest in performing a task. It is presumed that these reliable properties of the environment could be inferred from recurring properties that are learned by experience and over evolutionary time scales (Geisler and Diehl, 2002). Consequently, studying visual behavior using natural images is meaningful as it is most relevant to the circumstances that the visual system has been exposed to. In addition to statistical properties of the environment, characteristics of the visual system could also be constrained by particular tasks that are required to perform in order to survive. One of these important tasks is known as detection. Detection is a fundamental task of interest as it is relevant to performing more complex visual tasks that guide behavior. However, relatively little is known about our capabilities and the underlying neural mechanisms of detection under natural conditions. For instance, little is known about how humans and non-human primates detect a known target when it is added to a background composed of patterns actually observed in the natural world. This thesis builds on earlier findings in the vision literature to explore neurophysiological underpinnings of detection. Specifically, the main concern of this investigation is to improve our understanding of population responses in the primary visual cortex (V1) that are relevant to detection in natural backgrounds. One specific goal is to determine whether it is possible to quantitatively predict behavioral performance from the population responses to targets added to natural backgrounds. In the following section I address the significance of studying detection using natural stimuli, seeking an appropriate animal model, and exploring visual coding in V1 populations.

BACKGROUND AND SIGNIFICANCE

Detection is a fundamental visual task. In a typical detection task an observer is required to identify the presence of a target signal when added to a background (often called a mask). The challenge to the observer is to identify the trials in which the target is present in the background from those trials in which there is background alone. Varying the stimulus conditions of the target and background can make this a very challenging task for the visual system. Detection is fundamental to essentially every task that primates do as this capacity is required to identify signals that are relevant to perform a more complicated tasks such as visual search and object recognition. Nonetheless, most studies of detection use simple artificial backgrounds because they allow precise control over independent variables (Foley, 1994; Watson and Solomon, 1997; Goris et al. 2013). Traditionally, a majority of detection experiments with background masks rely on synthetic noise derived from stationary processes, such as white noise and 1/f noise. These simple backgrounds are attractive to use when formulating closed-form analyses of background masking effects, such as those provided by information theory (Shannon 1949) and signal detection theory (Green & Swets, 1966). For example, signal detection theory shows that the optimal filter to recover a target signal in white noise is a matched template filter (Green & Swets 1966; Eckstein et al. 1997). Another common choice of background noise is 1/f noise, mainly because its spectral structure is widely found in nature (the average amplitude spectrum of natural images falls as 1/f). While 1/f noise may be a better model stimulus for natural vision the phase structure in 1/f noise is random, like in white noise. This is not true for natural images that have considerable phase structure due to the presence, for example, of edges and contours induced by the presence of objects.

Although statistical properties of natural image backgrounds and artificial backgrounds are very different, analytically describing masking effects of natural images

is often considered intractable due to the large amount of variability in natural images. Nonetheless, there have been efforts to simulate and measure detection performance in natural backgrounds. A previous study (Bradley et al. 2014) used a wide range of conditions for target detection and conducted a series of behavioral experiments using various targets and multiple types of backgrounds—including natural scene backgrounds. In this study, a physiologically inspired model (Retina-V1 model) was capable of predicting human detection sensitivities of multiple targets placed across arbitrary locations in the visual field in arbitrary backgrounds. Specifically, the Retina-V1 model was designed by incorporating well-known physiological factors of the retina and V1, followed by pooling V1 responses in a near-optimal manner. Furthermore, successful model predictions of detection performance based on model responses of the retina and V1 responses motivate physiological assessment of target-detectability in V1 populations.

To access V1 activity, I used macaque monkeys as a surrogate animal model for studying the human visual system. Historically, macaques have been widely used as an alternative animal model to humans in exploring neural mechanisms due to similarities in the anatomy and physiology of the visual system. However, to the best of my knowledge, macaque detection performances in natural backgrounds have not been reported and it is unknown how capable macaques are in performing detection in natural backgrounds. This study evaluates macaque detection performances and compares this to human performances reported in the previous study by Bradley et al. (2014). This comparison provides a quantitative assessment of how well macaques qualify as an animal model for studying human detection performances and the underlying neural mechanisms that mediates detection.

From recordings of macaque V1, this study explores properties of neural activity relevant in identifying a target in natural scenes. Specifically, two important aspects are

taken into consideration when characterizing V1 physiology in response to natural scene stimuli. First, I seek to understand how V1 neurons work together in rendering a target pattern at the level of populations instead of monitoring individual recordings of a few cells. Second, I seek to understand whether visual coding is different between natural image stimuli and simple artificial stimuli.

The focus here is on neural activity at the level of populations was considered because complex natural stimuli are likely to be represented by populations rather than a few number of cells. The response of V1 neurons to simple stimuli, such as isolated bars or gratings presented within their receptive field, is often described as orientation and spatial frequency tuning (Hubel and Wiesel, 1959; DeValois et al., 1982). This response behavior shares many properties of linear filters that encode a visual input in terms of the magnitude of 2D Fourier components. However, such filter properties cannot fully account for a wide range of complex stimuli. For example, a V1 neuron's peak response to a preferred stimulus in orientation and spatial frequency can be suppressed or facilitated by additional stimuli placed either within or outside the classical receptive field (Gilbert and Wiesel, 1990; DeAngelis et al., 1992). This modulation can vary from one cell to another according to the relative amount of overlap between the compound stimulus and the cell's preference in orientation, spatial frequency, stimulus location, and contrast (De Valois and Tootell, 1983; Walker et al., 1999; Polat et al., 1998). In general, the response of a neuron to a compound stimulus could not be reliably predicted by building up from individual responses to simple stimuli. This provides evidence that coding of complex stimuli is difficult to characterize at the level of individual cells. Therefore, studying populations contributes to advances in our knowledge by testing whether distributed cortical activity conveys information that can better (and more parsimoniously) explain neurophysiological mechanisms for detection.

In studying populations, the functional architecture in V1 is taken into consideration, as V1 neurons are not randomly distributed but functionally organized in an orderly manner. V1 has a topographic map of the visual field that is referred to as the retinotopic map. The visual field is mapped with global distortions, which are particularly pronounced in humans and other primates where the central visual field is magnified at the expense of the periphery. From optical imaging studies we know that the retinotopic response to a small pattern stimulus elicits a cortical spread that reflects the stimulus location and shape (Chen et al. 2006; Michel et al. 2013). When background masking is minimal, as in uniform backgrounds, the retinotopic activation could be sufficient to distinguish the target signal from background noise in V1. On the other hand, it is unclear how retinotopic activation could overcome masking effects from the wide range of spatial features in natural backgrounds. Spatial features from natural backgrounds could elicit neural activities that could obscure target-evoked activities at the retinotopic scale.

When the magnitude of retinotopic responses is similar to the target and background, additional sources of information are necessary to distinguish the target from background. At the population level some of these additional sources of information may be carried in the columnar structure of V1. Particularly, in addition to the retinotopic map, V1 neurons in macaques are organized in columns according to various stimulus features such as orientation, ocular dominance, direction, and spatial frequency (Nauhaus et al., 2012). As an exploratory attempt, this thesis focuses only on orientation column activities, as orientation selectivity is a prominent feature in most V1 neurons. The functional architecture of orientation columns is striking in their organization and precision but their significance in conveying information to perform a task such as detection is unknown. As V1 is the initial stage of the visual hierarchy in the cortex (Felleman and Van Essen, 1991), I asked whether downstream areas of the visual hierarchy may utilize population responses

from orientation columns to convey information of a target pattern in natural backgrounds. By comparing target detectability from orientation columns to behavioral detection performances, this study could provide some insight to this question.

Aside from the significance of analyzing neural activity in populations, another important aspect is considered when using natural images. Visual coding may be different between natural scenes and artificial stimuli. Particularly, it is unclear how stimulus-tuning constancy holds in natural scenes where various spatial features are present at the same time. For instance, in the case of encoding orientation information, contrast-gain control is a well-known mechanism that maintains orientation tuning across a wide range of contrasts. Neurons in the early stages of the visual pathway adjust their range of representing luminance contrast through a divisive gain-control operation in the dimension of contrast (Albrecht & Geisler 1991; Heeger 1991, 1992; Geisler and Albrecht, 1992). The implication of contrast-gain control is to sacrifice accurate representation of contrast to render orientation tuning to be largely contrast invariant (Geisler and Albrecht, 1995; Finn et al., 2007). There are additional cases that reflect changes of the response gain in orientation tuning. One example is observed when spiking activity from a preferred orientation grating was shown to attenuate when an orthogonal orientation (cross-orientation) grating was superimposed (Morrone et al., 1982, Priebe and Ferster, 2006). Another case is surround suppression where areas of visual areas surrounding the classical receptive field reduce the responses of orientation-tuned neurons (Born and Tootell, 1991; Knierim and Van Essen, 1992) and especially for backgrounds in the preferred orientation (Cavanaugh, Bair, and Movshon, 2002). This implies that in the presence of a wide range of orientations, as in natural images, the responsiveness to a specific orientation would reduce. A simple account of this observation poses a challenge to the visual system in faithfully capturing various spatial structures that are composed of orientation edges.

However, perhaps the visual system has adapted a way to circumvent this challenge by recruiting contextual information from the surroundings of a cell's receptive field. Surrounding contextual influences in V1 neurons have been reported where similar structures surrounding a cell's receptive field recruited stronger surround suppression relative to heterogeneous surroundings in natural images (Coen-Cagli et al. 2015). Computationally, suppressing redundant patterns could help segregating visual information from a heterogeneous pattern from its surrounding.

In previous paragraphs, the potential importance of using natural backgrounds in studying detection and its underlying mechanisms was addressed. A particularly motivating factor for the current study was the successful modeling reported in a previous study that used well-known physiological factors to predict human detection performance across various stimulus conditions, including natural scene backgrounds (Bradley et al. 2014). The current study measured neural activity in V1 and compared that activity to behavioral detection performances. Macaques were used as an animal model and visual coding in V1 populations was measured at the scale of the retinotopic map and orientation columns. By analyzing behavioral detection performances and V1 activity in the macaque, this study provides new insight to pooling strategies in V1 populations that are predictive to detection performances.

HYPOTHESIS

The main hypotheses is that detection performances in natural backgrounds can be predicted from the target detectability of pooled responses at the scale of the retinotopic map and/or from the orientation-column map. The performance measures were psychometric functions estimated from the behavioral responses and the neurometric functions estimated from the neural responses at the retinotopic and orientation-column

scales. The relationship between behavioral thresholds and neural population thresholds estimated from psychometric and neurometric functions were compared to test the hypothesis.

SUMMARY OF FINDINGS

The results indicate that macaque thresholds for detection in natural backgrounds are similar to human thresholds (up to a scale factor). Furthermore, the pattern of behavioral performance in the macaques was predicted from the neural signals both at the scale of the retinotopic map and at the scale of the orientation columns, using different pooling strategies in the two cases. These findings suggest that signals at both scales are related to detection performance, but with a relatively better correspondence for the orientation-column responses at higher background contrasts. At higher background contrasts, it could be the case that orientation columns convey neural signals that are more reliable for downstream areas in detecting a target in complex environments. Hopefully, findings from this study can be used in future studies to relate V1 population responses with behavioral choices for detection in natural backgrounds.

Chapter 2: Behavioral Experiment

Animal experiments, particularly with macaques, have contributed much to our understanding of visual mechanisms. The macaque visual system is often used as a model for the human visual system, where it is possible to apply highly informative invasive techniques such as single unit recordings, neural tracers, micro-stimulation, histology, and optical recording. In general, macaque and human brains are similar in early stages of the visual hierarchy—especially in V1, V2, V4, and V5/MT—but differ in several higher-order cortical regions (Tootell et al., 2003). Homologous regions in the visual system motivate the use of macaques. However, this level of homology between macaques and humans does not guarantee that macaques can be used to predict human behavioral or neural detection performance. Behavioral detection performance in natural scenes has not been compared between the two species and it is therefore unknown how macaques qualify in studying human detection performance. The focus of this chapter is to investigate the extent of similarity and disparity in detection performance, in order to test whether macaques are a viable animal model for human observers in natural detection tasks.

STIMULI

A previous study provided human psychophysical measurements in detection across various background types, targets, and eccentricities (Bradley et al., 2014). A subset of experimental conditions from the human study was replicated to measure macaque detection performances.

The stimulus consisted of a natural background patch and an additive Gabor target that was present in 50% of the trials. Natural images were randomly selected from a set of 100 natural scene images from the same image database (Geisler & Perry, 2011) used in the human study and were converted to an 8-bit grayscale. These images were

"Gaussianized" by matching their pixel histograms to a 1/f noise image (Figure 1A). This procedure effectively preserved the spatial frequency, orientation, and phase structure of natural images while matching the mean luminance and contrast of 1/f noise images. This allowed Bradley et al. to compare masking effects from natural images and 1/f noise images. In this thesis, only Gaussianized natural images were used in order to compare macaque and human performances in natural backgrounds. Similar to the human study, the background stimulus was constructed by cropping a 4o image patch from the Gaussianized natural image. A raised cosine windowing function was applied to the cropped image patches to smooth image boundaries (Figure 1B, left). Contrast values were approximately the same for the larger images and cropped patches.

On half of all trials, a horizontal Gabor (sine-wave grating multiplied by a 2D Gaussian envelope) was added to the center of the background patch (Figure 1C, right). Dimensions of the Gabor target were matched to the human study: 0.84o in diameter, 4 cyc/deg in spatial frequency (1 octave bandwidth), ~2.5o eccentricity from the center of the fovea.

I varied the contrasts of both the background and the target. Background images were adjusted to have a constant mean luminance of 30 cd/m2 and the RMS contrasts varied across four levels: 0%, 1.875%, 3.75%, and 7.5%. At each level of background contrast, target contrast was varied to measure psychometric functions from three macaques (monkeys C, H, and T). Although the experimental paradigm and stimuli were closely matched to the human study by Bradley et al. (2014), background contrasts in this study did not cover the full range of the human study. In comparison to macaques, it was possible to measure human detection thresholds using a lower range of target contrasts. Human observers reached perfect performances at relatively lower target contrasts for all levels of background contrast and hence psychometric functions could be measured at

higher background contrasts. However, this was not the case with macaques. Macaque detection thresholds were found to be higher (results) which required a higher target contrasts to reliably estimate psychometric functions. Pixel gray-level clipping occurred when trying to measure macaque thresholds for background contrasts exceeding 7.5%.

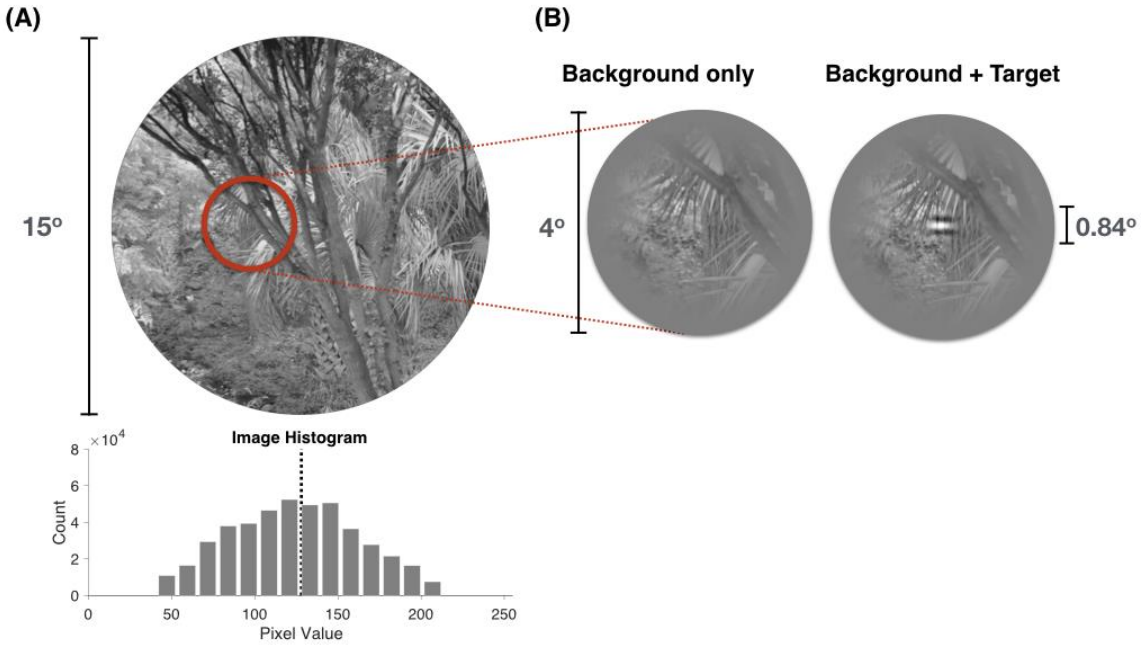

Figure 1: **Natural image background and target stimulus.** (A) Gaussianized natural scene image: an example of a "Gaussianized" natural scene image (15% contrast) used in both human and monkey experiments. The original collection of natural scene images was provided by Geisler & Perry (2011). A large image was converted to an 8-bit image and the contrast was set to 15% contrast (RMS). The pixel histogram was adjusted to have a Gaussian distribution that matched a 1/f image of the same size (not shown). Below the image the histogram of pixel intensities is shown. The mean luminance value was set at 128 (dotted line). (B) *Left*, background-only stimulus. The background stimulus is a 4o patch that is cropped from (A) with a raised cosine envelope applied to the cropped image. *Right*, target is added to background at 14% contrast (RMS). A horizontal Gabor target (0.84° diameter, $\sigma = 0.14^{o}$, 4 cpd) is added to the center of the background stimulus.

DETECTION TASK

The task for both humans and macaques was a single-interval forced-choice task (yes/no) where a target was added to a natural image background on half of all trials. Similar to the human study, contrast detection thresholds were measured in macaques from natural backgrounds that were presented at various contrast levels (1.875%, 3.75%, and 7.5% background RMS contrast).

Three male macaques (monkeys C, H, and T) were trained to perform a reaction-time task to detect a small horizontal Gabor target (Figure 2). Each trial began when the monkey successfully maintained fixation on a uniform gray background on a CRT display. Initial fixation durations randomly varied from 1000 *ms* to 1300 *ms*. At the beginning of the initial fixation, a brief audible cue was presented with the fixation point. Initial fixation was used to ensure that the animal was fully engaged in the task. Following initial fixation, the fixation point dimmed and the stimulus was presented after 250 *ms*. When the target was absent, the monkey was required to maintain fixation within a small window (<1o full width) for 1 second to receive a liquid reward. In target-present trials, the monkey was required to saccade to the target location within 600 *ms* and maintain gaze at that location for an additional 300 *ms* to receive the reward. Target-present and target-absent trials were randomly interleaved and the duration of each trial lasted from 2 to 3 seconds, with a fixed inter-trial interval of 2.5 seconds. Audible feedback was given at the end of the trial. A liquid reward accompanied the audible feedback for a positive feedback for hits and correct-rejections. For misses and false-alarms, an extra 3 second interval was added to a different audible feedback.

Eye movements were measured using an infrared eye-tracking device (Dr. Bouis Inc.). If the eye position deviated by more than 0.5o from the fixation point, the trial was discarded and another trial was added to the block. Monkey reaction times were defined as

the duration from stimulus onset until the time of saccade initiation towards the target location. Reaction times were systematically shorter for higher target contrast and longer for lower contrasts. The median and mean reaction time was approximately 200ms (Figure 3).

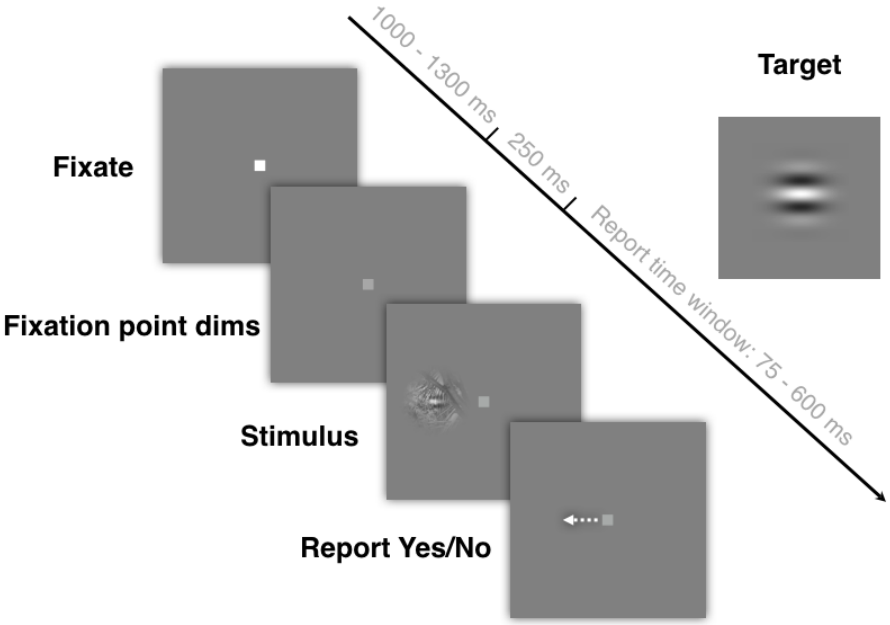

Figure 2: **Detection task sequence.** The task for both humans and macaques was a single-interval forced-choice task where the target was present on half of all trials. The stimulus consisted of a patch of a natural background and a Gabor target ($\sigma = 0.14^{o}$, $\theta = 0^{o}$, 4 cpd) shown in the upper right. For a set of background images at a particular contrast, the target contrast was varied to measure a psychometric curve.

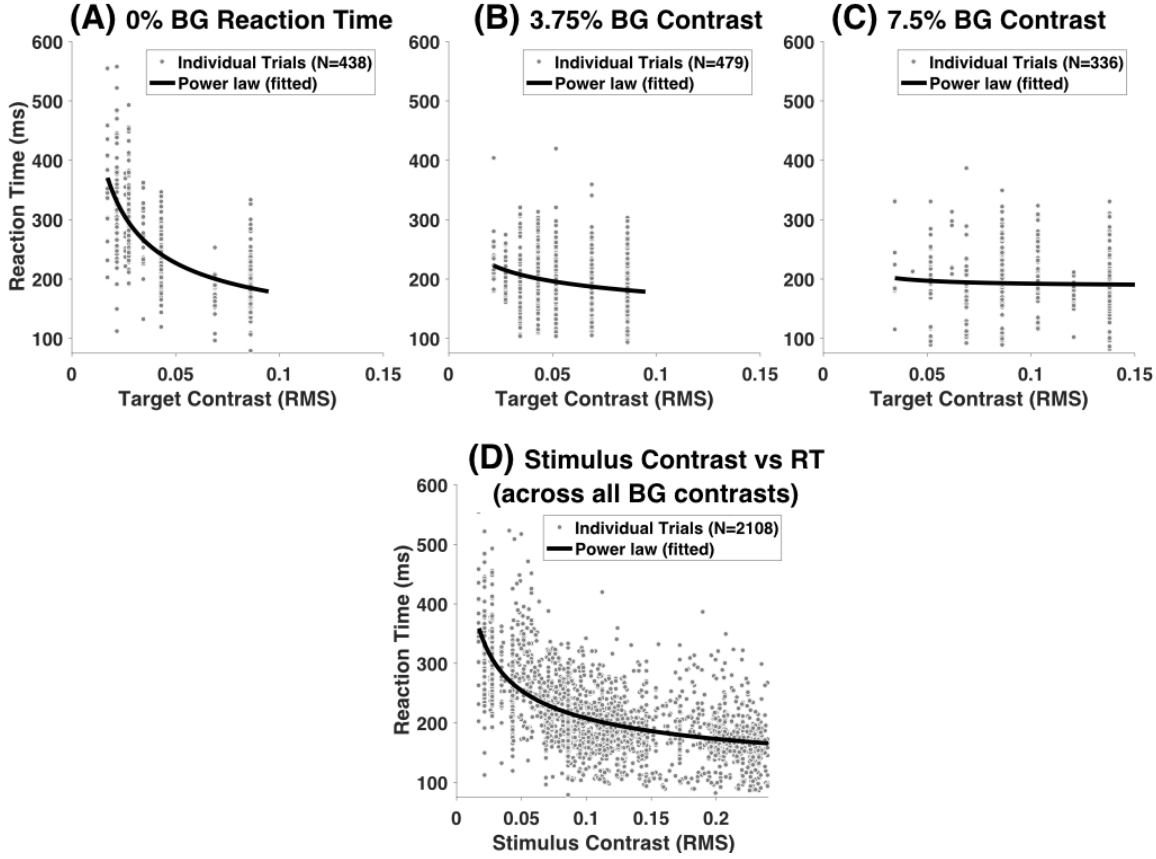

Figure 3: **Macaque reaction time**. Reaction times from macaques C, H, and T are plotted. For a successful trial when the target is present (hit trial), the reaction time is defined as the duration from stimulus onset to saccade initiation towards the target location. (A, B, and C) Reaction times are plotted separately for different background contrasts at 0%, 3.75%, and 7.5% as a function of the target contrast. (D) Reaction times are plotted as a function of stimulus contrast, which is the overall RMS contrast of the target and background under the target region. Dots correspond to a trial's reaction time and the solid line is a decaying power function that was fitted to the data to visualize the overall trend. The monkey was trained to maintain fixation for at least 75 *ms* after target onset before making a saccade and was required to make a saccade before 600 *ms* after target onset to be categorized as a hit.

ANALYSIS

Psychometric functions were measured in a single-interval, forced-choice paradigm where the target contrast was blocked. At each background contrast, a psychometric

function was measured based on at least 200 trials. For a given background contrast, five blocks of at least 40 trials each were run in random order of target contrast. For each block, there were two conditions (target-absent & target-present) and four types of outcomes (hit, miss, false-alarm, correct-rejection). The observer's performance can be reported in terms of the percentage of correct response or in terms of probabilities for each possible outcome. Although there are four types of outcomes, one does not need four numbers to summarize the behavioral performance; hits and false-alarms are sufficient to describe the observer's behavior.

Although the observer's behavior can be reported as the proportion of each outcome, this is not completely satisfactory in providing an unambiguous measure of the observer's detectability. A bias-corrected summary statistic that represents the observer's detectability is preferred to compare across stimulus conditions and observers. The quantity, d', is derived from signal detection theory (Green and Swets, 1966; Wickens, 2002). The foundation is drawn from statistical decision theory and is similar to the ideas that are used in statistical testing to make a decision between two hypotheses (e.g. target-absent vs. target-present). The idea here is to consider an internal response of the observer that can be represented by a point along an underlying decision axis. I assume that internal responses of target-absent and target-present stimuli follow a conventional Gaussian distribution that represents both external and internal noise (variance). Moreover, as the target signal is additive, the two distribution can be assumed to have equal standard deviation. Under these conditions, target-absent trials are samples from the noise distribution and target-present trials are samples from the signal-plus-noise distribution. The signal-to-noise ratio (SNR, or commonly, *d'*) is then computed by taking the difference between the two means and dividing it by the standard deviation. When considering Gaussian distributions of equal variance, it has been shown that an observer's percentage

of correct responses is related to d' values using the normal cumulative distribution function, Φ (Green and Swets, 1966; Bradley et al., 2014):

$$\%correct = \Phi\left(\frac{d'}{2}\right) \tag{1}$$

In simple detection tasks, the detectability of the target is varied by varying the target contrast. Thus, the ratio between the target contrast (c) and threshold contrast (c_{th}) determines the signal-to-noise ratio. Here the threshold is defined as the target contrast giving a signal-to-noise ratio of 1.0 (d' = 1.0). An additional parameter, beta, is required to describe the steepness of the psychometric function:

$$d'(c) = \left(\frac{c}{c_{th}}\right)^{\beta} \tag{2}$$

However, notice that the threshold is independent of the value of the steepness parameter, β. The basic form of the predicted psychometric function, assuming there is no bias in the monkey's decisions and no lapse rate, is derived when we map d' values into success rates (percent-correct):

$$\%correct = \Phi\left(\frac{1}{2}d'\right) = \Phi\left(\frac{1}{2}d'(c)\right) = \Phi\left(\frac{1}{2}\left(\frac{c}{c_{th}}\right)^{\beta}\right) \tag{3}$$

However, I found that the monkeys were sometime biased (γ) and had a lapse rate (λ), thus better fits were obtained by including parameters for these two factors. Specifically, the equations for the false alarm and hit rates were given by the following two equations, which are a generalization of equation (3):

$$P_h = \lambda + (1 - 2\lambda)\Phi\left(\frac{1}{2}\left(\frac{c}{c_{th}}\right)^{\beta} - \gamma\right)$$

$$P_{fa} = \lambda + (1 - 2\lambda)\Phi\left(-\frac{1}{2}\left(\frac{c}{c_{th}}\right)^{\beta} - \gamma\right) \tag{4}$$

This pair of equations has four parameters: contrast threshold (c_{th}), slope (β), criterion bias (γ), and lapse rate (λ). Maximum likelihood estimation (MLE) was used to estimate the parameter of equations (4) from the behavioral responses from each monkey. Specifically, I maximized the log likelihood of all the responses from a given condition:

$$\log L(c_{th}, \beta, \gamma, \lambda) =$$

$$\sum_{i=1}^{n} N_h(c_i)\ln P_h(c_i) + N_m(c_i)\ln (1 - P_h(c_i)) + N_{fa}(c_i)\ln P_{fa}(c_i) + N_{cr}(c_i)\ln (1 - P_{fa}(c_i)) \tag{5}$$

where ***n*** is the number of contrast levels of the target, and $N_h(c_i)$, $N_m(c_i)$, $N_{fa}(c_i)$, and $N_{cr}(c_i)$ are the numbers of hits, misses, false alarms, and correct rejections, for contrast level c_i.

RESULTS

Behavioral data were collected from three macaques (C, H, and T) across four levels of background contrasts (0%, 1.875%, 3.75%, and 7.5%). Maximum likelihood fits of equation (5) to behavioral data provided estimates of the detection threshold, steepness (slope), lapse rate, and criterion bias for each condition. Figure 4A plots contrast detection thresholds for the macaques and humans (from Bradley et al. 2014). In Figure 4A the macaque thresholds are represented in red circles and the red line corresponds to Weber's law for contrast power that was fitted to the data. Weber's law for contrast power indicates that the threshold contrast power is proportional to the background contrast power. This relationship appears curved at low contrasts in the figure, because the axes are in units of contrast rather than contrast power. Human thresholds from Bradley et al. are shown in gray circles. Weber's law for contrast power in human data is plotted as well. Figure 4A suggests that threshold contrast power increases in a similar fashion as a function of

background contrast in both macaques and humans, but with lower thresholds (higher sensitivities) for human observers. Since Weber's law for contrast threshold hold in both species, the relative trend in performances might reveal a tighter relationship between macaques and humans. To better compare the relative change in performance, thresholds were normalized by the first data point from each plot (the threshold on a uniform background) and replotted in Figure 4B. In Figure 4B, normalized macaque thresholds overlap with the relative trend of human thresholds (Weber's law for contrast power). From this analysis we see that there is a relatively good agreement between the relative contrast thresholds in macaques and humans for natural backgrounds.

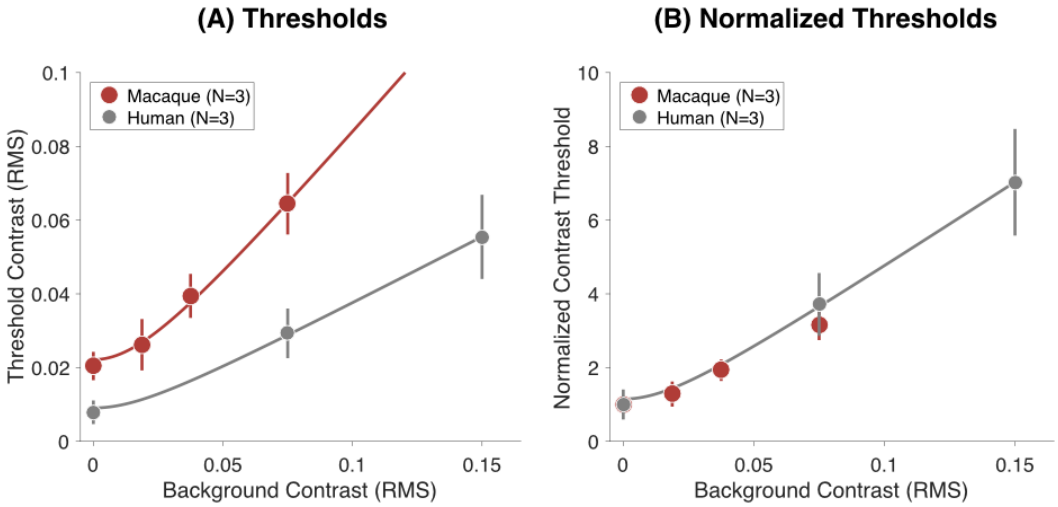

Figure 4: **Detection sensitivities between macaques and humans**. Three macaques were trained to perform a detection task using the same stimuli in the human experiment. Behavioral data from macaques were incorporated only after the monkey was sufficiently trained. (A) Contrast detection thresholds are plotted in circles for macaques (*red*) and humans (*gray*). Weber's law for contrast power is plotted in solid lines for each species. (B) The relative change in detection thresholds is shown. Each threshold value was normalized by the threshold from a uniform blank background. The trend in contrast masking shows a close resemblance between macaques and humans. Error bars correspond to the 95% confidence interval.

Although the relative trend of thresholds is similar, slopes of the psychometric functions were different between the two species. From the human study of detection, the slope (or steepness) of the psychometric functions were fairly constant over all conditions. In macaques, slopes were significantly higher across all background contrasts, and decreased slightly with background contrast (Figure 5A). One potential insight relevant to this observation is provided by a past study in contrast detection (Pelli, 1985). In this study, elevated thresholds that were accompanied with steeper slopes were noted to be consistent with an overall increase in stimulus uncertainty. The source of uncertainty could potentially be extrinsic (in the stimulus) or intrinsic (in the observer). Given that the current stimuli were nearly identical between macaques and humans, the results suggest that under the current conditions macaques may have a higher level of intrinsic uncertainty than humans.

A systematic trend in macaques was observed in the lapse rate parameter, as shown in Figure 5B. As contrast masking increases, there was a substantial increase in the lapse rate. The standard interpretation of the lapse rate is that there is some random proportion of trials where the subject is not paying attention or produces a random motor error. The simplest version of this hypothesis would suggest that the lapse rate should be relatively constant across background contrasts, rather than showing a systematic trend. Therefore, something more complicated appears to be going on. In the case of human observers, a lapse rate was not observed or was small enough to be inconsequential in fitting the data.

Lastly, as mentioned before, an observer's performance can be affected by a criterion bias in forced-choice tasks. To correct for such effects, a criterion parameter was incorporated for bias correction (equation 4). Criterion bias is reported in units of d' values in Figure 5C. On average, macaques were slightly more conservative than human subjects.

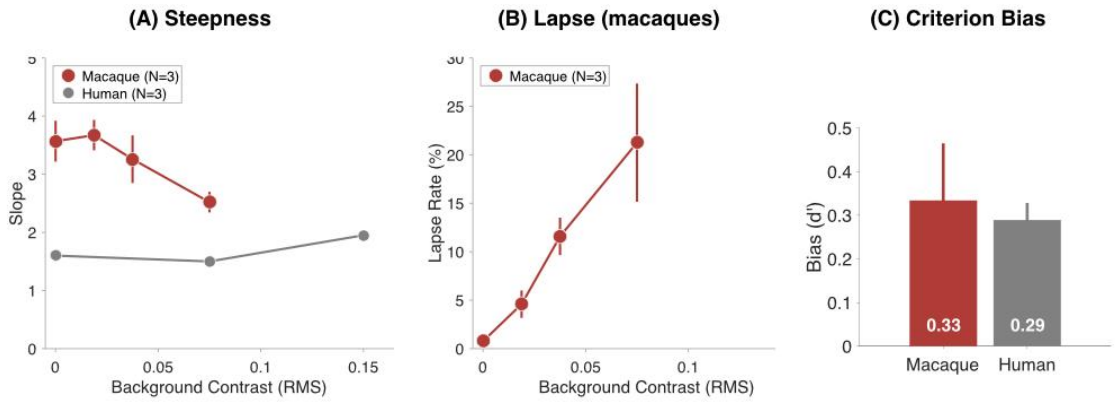

Figure 5: **Maximum likelihood estimates of the psychometric function slope, lapse rate, and criterion bias**. (A) The slope of the psychometric function at each background is plotted. Human performances show stable slopes across conditions whereas macaque slopes were much higher on average. (B) Lapse rate is reported for macaques only. Human observers did not show any lapse and therefore is not reported here. (C) Criterion bias is plotted in units of d' for both species. Error bars correspond to the 95% confidence interval.

In summary, macaque detection performance was measured for a subset of the natural images used in the human study by Bradley et al. (2014). Weber's law for contrast threshold holds for both species but with higher thresholds in macaques. Moreover, in the case of macaques, the slopes from psychometric functions were higher and the lapse rate increased as a function of background contrast. However, the critical finding concerns from the similarity between the relative detection thresholds across the two species. As background contrast increased, the relative change in performance was nearly identical for macaques and humans. This suggests that macaques provide a useful model for gaining insight into the neural mechanisms underlying human detection performances in natural backgrounds.

Chapter 3: Physiology

In the previous chapter, macaque detection sensitivities were measured using a simple detection task with natural scene backgrounds. Behavioral results indicate that detection performance of macaques and humans change in a similar manner. The relative trend of detection sensitivities between macaques and humans were in good agreement but with lower sensitivities in macaques. Similarities in detection performance and homologies in the visual system suggests that the macaque is an informative animal model of the human visual system in studying detection in natural scenes. Therefore, for practical reasons, macaques were used in recording cortical activity while the monkey viewed a target in natural scene backgrounds.

The main aim of this chapter is to explore neural activities that are relevant to identifying a target in natural scene backgrounds. Specifically, I would like to describe how V1 population responses can be pooled to predict detection performances in natural backgrounds. From a previous study by Bradley et al. (2014), human detection performances could be predicted by modeling well-known properties of the retina and V1. In a separate study using macaques, V1 responses could be optimally combined to outperform target detection performances in a uniform blank background (Chen et al., 2006). These findings provide motivating factors to investigate whether it is possible to quantitatively predict behavioral performance from V1 population responses to a fixed target added to natural scene backgrounds.

Overall, this chapter aims to find a correspondence between physiology and behavior by assessing how V1 population activities can predict psychophysical measurements from Chapter 2. As an initial attempt, population responses are recorded from a fixating monkey with VSDI while the visual stimulus is presented near the fovea

(1.5º eccentricity). Target detectability in population responses are estimated at the scale of retinotopic signals and orientation column maps, and benchmarked to behavioral detection sensitivities.

First, stimulus-evoked responses in the retinotopy are assessed as the topographic layout of the visual field is a prominent feature of V1. For a small Gabor target on a blank screen, the retinotopic response resembles a two-dimensional Gaussian envelope when imaging from superficial layers of V1, as shown in the left columns of Figure 9A&B. When using similar Gabor stimuli, retinotopic responses were shown to be sensitive in rendering the location and shape of a target on uniform backgrounds (Chen et al. 2006; Michel et al. 2013). However, relatively little is known how retinotopic signals contribute to target detection in the presence of natural background masking. In uniform blank backgrounds, the two-dimensional Gaussian response pattern could provide a reliable source of information in identifying the target (Chen et al., 2006). On the other hand, when background noise is strong, the location and shape of the retinotopic response pattern could be obscured by a wide range of spatial structures in the background that can evoke V1 populations in a similar manner. It is unknown whether such cases provide a factor in decreasing detectability. To approach this question, the optimal decoding model from Chen et al. (2006) is applied to retinotopic responses of a Gabor target in natural scene backgrounds. The rationale for using an optimal decoding model is to determine how well the best possible pooling could do. The performance from this model is based on the optimal use of information available in the retinotopic scale, given the limits imposed by the recording technique and neural noise. The optimal decoder is not a description of how the brain combines retinotopic signals but is useful in providing the standard needed to make a comparison with behavioral performance. In essence, the comparative analysis is interesting, because deviations between model performances and

behavioral performances are likely to reveal insights about the actual neural process underlying the task.

Second, orientation column responses are also analyzed in comparison to behavioral performance. As a wide range of spatial structures are present in the natural environment, I hypothesize that orientation features could provide a complementary source of information in segregating a target pattern from natural backgrounds. In this case, the structure of the target could be conveyed in orientation-selective cells that reliably capture features of the target pattern in natural background noise. To quantify the extent of target-detectability in orientation columns, a matched-template decoder is applied to orientation column responses. The columnar response to the target stimulus is used as a template to derive a decision variable for each trial condition and to ultimately calculate the signal-to-noise ratio between target-present and target-absent conditions. Unlike the optimal decoder used for retinotopic signals, optimal performance is not guaranteed since the statistical structure of VSDI responses of columnar responses are unknown. Nonetheless, the matched-template model provides a tractable approach as the template is linearly weighted to each trial condition. In comparison to retinotopic signals, orientation column responses are analyzed to explore a different source of information that is encoded in V1 populations. A different (and potentially sub-optimal) read-out method is applied to orientation columns and the resulting target-detectability is compared to behavioral performance.

In summary, voltage-sensitive dye imaging (VSDI) recorded V1 population activity on fixating macaques using the same stimuli from the behavioral experiment. Neural sensitivities of V1 populations at the scale of the retinotopic map and orientation columns are analyzed with separate candidate pooling methods. Thereafter, neural sensitivities are benchmarked to behavioral detection sensitivities to relate potential read-

out methods of V1 population responses that are predictive of behavioral detection performances.

VSD IMAGING

Wide-field imaging with voltage-sensitive dyes (VSDs) is used to record neural population activities at a high resolution in space and time (Grinvald et al., 1988; Shoham et al., 1999). Before each imaging experiment, VSDs are topically applied to the cortex which bind to neuronal membranes. VSD molecules are activated by light at a wavelength of ~630 *nm* and fluorescence from neuronal activity are reflected back in the range of infrared wavelengths that are selectively filtered by a dichroic mirror. VSD molecules produce an instantaneous and linear change in fluorescence in response to membrane potential activity. Responses from VSD reflect a local sum of sub-threshold neural activity from the spread of dendrites and axons at the superficial layers of the cortex (Grinvald and Hildesheim, 2004).

In this thesis, VSDs were applied to macaque V1 through a surgically implanted recording chamber. After staining, fluorescence from neuronal activity was recorded through an imaging data acquisition system (Optical Imaging, Inc.). The imaging system was configured to record a cortical region of approximately 8 x 8 mm (Figure 6B). Imaging data were collected at 110 Hz where each frame was consisted of 512 x 512 pixels. Subsequent analyses at the scale of retinotopic and orientation columns were based on VSD responses that were spatially binned to a resolution of 64 x 64 pixels where each binned pixel, or site, corresponds to 0.11 x 0.11 *mm* in the cortex. Spatial binning was necessary to filter out high frequency shot noise and subsequent analyses were performed on the binned data which effectively suppressed high-frequency noise (Chen et al., 2006; Chen et al., 2012). Further details of experimental techniques for optical imaging with VSD in

behaving monkeys have been described elsewhere (Seidemann et al., 2002; Slovin et al., 2002; Arieli et al., 2002).

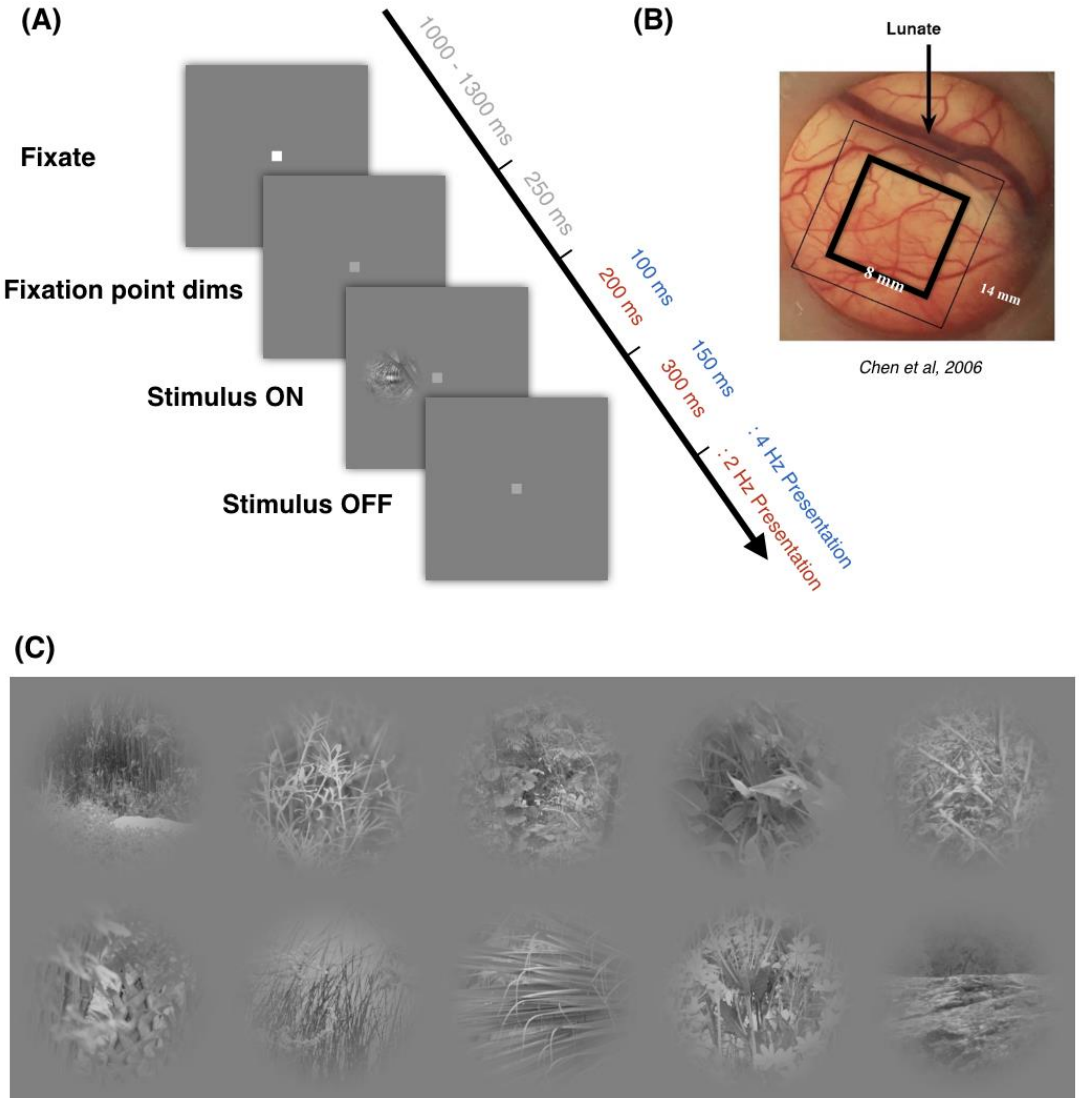

Figure 6: **Outline of the imaging experiment**. **(A) Task Sequence**: VSDI recordings were done on fixating monkeys. Stimuli were presented in two different temporal profiles. For monkey B, the stimulus was on for 100 *ms* and off for 150 *ms* for a single presentation, and repeated four times during 1 second (*blue*). For monkey A, the stimulus was on for 200 *ms* and off for 300 *ms* , and was repeated twice within 1 second (*red*). Although stimulus presentation durations differed, the same integration time interval was used (150 *ms*). **(B) Cranial window**: The imaging region was configured to record the inner black square with thicker borders (8 x 8 mm^2; adapted from Chen et al., 2006). **(C) Background stimuli**: Ten image patches used in the imaging experiment are listed. Patches were manually selected from Bradley et al. (2014) to cover a range of distinct spatial properties.

FIXATION TASK AND STIMULUS

VSDI data were collected on fixating monkeys. Each imaging trial began with an audible cue and a small fixation point presented on a uniform blank screen. The animal was required to initially hold fixation in a range of 1000 to 1300 *ms* to ensure that the animal was fully engaged in the task. Following initial fixation, the fixation point dimmed to indicate the beginning of the trial and the stimulus was repeatedly presented during one second while the macaque maintained gaze around the dimmed fixation point.

Stimulus conditions varied by the background and target contrast, which follows the same specifications from the behavioral experiment (Chapter 2). Background images were adjusted to have a fixed mean luminance (30 cd/m^2) and the RMS contrasts varied across four levels (0%, 1.875%, 3.75%, and 7.5%) for monkey 1, and three levels (0%, 3.75%, and 7.5%) for monkey 2. At each level of background contrast, Gabor target contrasts varied from 1% to 14% RMS contrast.

Although stimulus conditions were the same with the behavioral experiment, a smaller subset of ten natural background patches were used in the imaging experiment. Image patches were manually selected to accommodate a range of distinct spatial structures in natural environments (e.g. low contrast, high contrast, dense, sparse, oblique, horizontal, and vertical structures). Similar to the behavioral experiment, the contrast of background patches was varied to adjust the level of contrast masking. A raised cosine window (4^o diameter at a viewing distance of 42.5 inches) was applied to all ten patches.

In terms of stimulus presentation, the presentation duration was different between two monkeys. For monkey 1, the stimulus was on for 100 *ms* and off for 150 *ms*, and repeated four times during one second (Figure 6A, *blue*). For monkey 2, the stimulus was on for 200 ms and off for 300 ms and repeated twice for one second (Figure 6A, *red*).

Although the duration of the stimulus presentation was different, neural responses were integrated using the same time window. For both monkeys, neural responses were integrated over 150 ms starting from the point where cortical activities began to rise in response to the stimulus. The first type of stimulus presentation will be referred as the '100 *ms* stimulus' and the second type will be referred as the '200 *ms* stimulus'.

ANALYSIS OF VSDI DATA

VSDI experiments were conducted in two monkeys (monkey B and A) that were not used in the behavioral experiment and only trained to perform a fixation task. Fifteen imaging sessions were collected from monkey B, and 10 sessions from monkey A. Each imaging session was analyzed separately as the staining quality differed for each experiment. Similar amounts and concentrations of VSD was applied for each imaging session. Although the VSD was controlled across experiments, staining qualities could fluctuate as the level of dye penetration varied depending on the condition of the cortical tissue. For this reason, summary statistics were computed for each imaging session and the grand average is reported at the end.

Data preprocessing was first applied to remove or suppress corrupt recordings. For example, trials with aberrant VSD responses were removed (< 5% of all trials). A trial was removed when the response amplitude was five standard deviations away from the average response amplitude of the entire imaging session. Furthermore, non-responsive cortical locations were suppressed by a spatial reliability map. The reliability at each cortical site (pixel location) was calculated by taking the reciprocal of the variance in fluorescence across time and across all conditions. In most cases, blood vessels and poorly stained locations produced higher variance in fluorescence that was independent of the stimulus condition and was weighted less with the reliability map. The reliability map effectively

suppressed unreliable regions from degrading the efficiency of pooling cortical responses across space.

VSDI analysis for retinotopic responses	
input	Imaging data across time and trials, ***I(x,y,time,trials)*** *e.g.* Raw : 512 x 512 pixels x 150 frames x 160 trials Binned : 64 x 64 pixels x 150 frames x 160 trials
process	**for each** *trial* **for each** *cortical site (x,y)* 1. ***F*** ← initial(maximum) fluorescence (0-100 ms) 2. ***F(t)/F*** ← normalize entire time course by ***F*** 3. Δ***F/F*** ← subtract blank conditions from all trials 4. **for the first** stimulus presentation cycle 4a. isolate response time course from single cycle 4b. anchor time course by pre-stimulus response **return** anchored initial response cycle 5. ***Shift integration time interval (150 ms)*** 5a. Uniform background only: **for each** target contrast condition * shift 150 ms interval with response latency * save integrated blank responses 5b. Uniform background + target: 5c. Natural background only: 5d. Natural background + target: ***** shift 150 ms interval with response latency 6. ***R(x,y,trial)*** ← integrate responses
output	Integrated neural responses, ***R(x,y,trial)*** *e.g.* Raw : 512 x 512 pixels x 160 trials Binned : 64 x 64 pixels x 160 trials

Table 1: Pseudocode for analyzing retinotopic responses

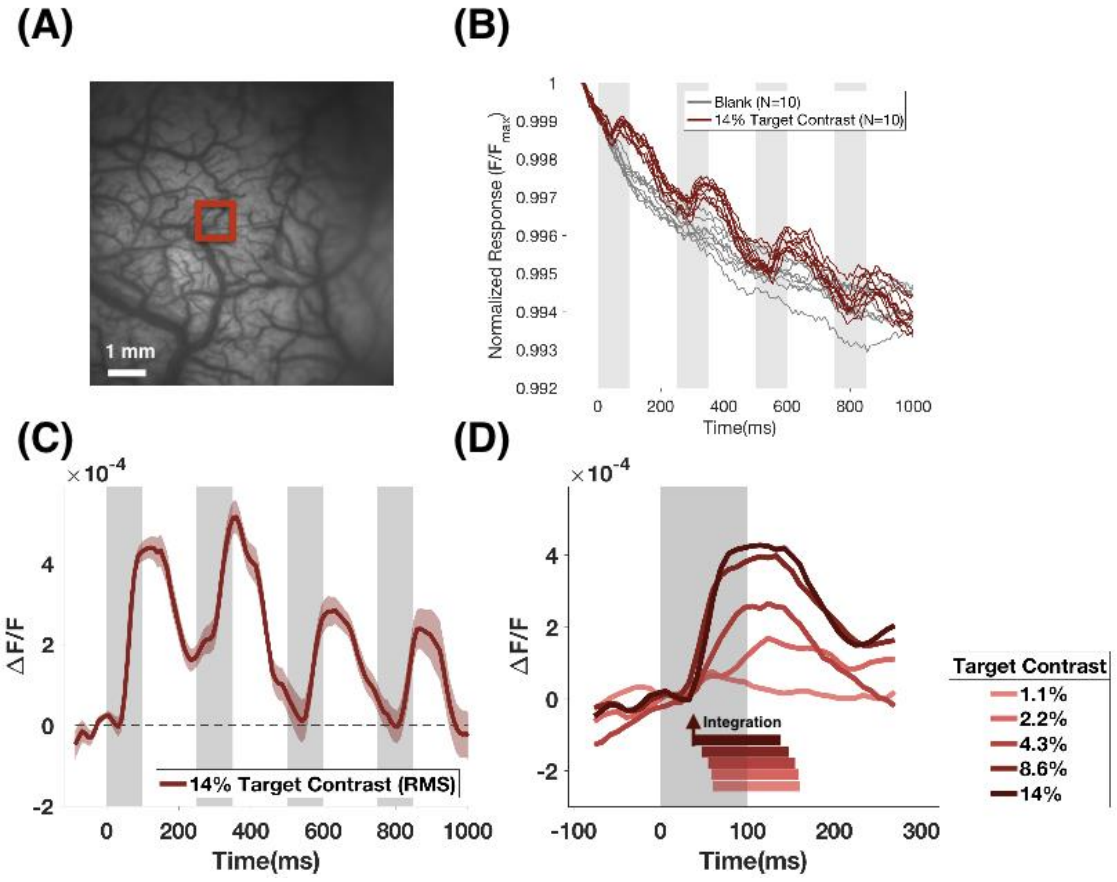
(A)
1 mm
(B)
Blank (N=10)
14% Target Contrast (N=10)
Normalized Response (F/F max)
Time(ms)
(C)
×10^-4
ΔF/F
14% Target Contrast (RMS)
Time(ms)
(D)
×10^-4
ΔF/F
Integration
Time(ms)
Target Contrast
1.1%
2.2%
4.3%
8.6%
14%

Figure 7: **Example VSDI time course from 100 ms stimulus**. An example dataset from monkey B (experiment date: 11/26/2014) is shown to illustrate how time courses of stimulus-evoked response were derived. Monkey B fixated at the center of the monitor while a Gabor target (0.84° diameter) was presented throughout four cycles (single cycle: 100 *ms* on, 150 *ms* off) on a blank screen. **(A) Example cortical site**: Subsequent time courses of VSDI responses correspond to the cortical region outlined by the 1 mm2 red square. This region corresponds to the center region of the retinotopic response to the Gabor stimulus. **(B) Normalized time courses**: VSDI recordings are normalized by the average response from the initial 100 ms of recordings (pre-stimulus). Red lines represent normalized time courses from trials that had a 14% RMS contrast Gabor target on a uniform background. The stimulus was shown four times (gray regions in the background). **(C) Blank-subtracted time courses**: Mean and standard error of the time course from (B) is shown after blank-subtraction. Gray regions correspond to time intervals when the stimulus is on. **(D) Response profile across target contrasts**: Average time courses from the initial stimulus presentation cycle is shown across target contrasts. Ten trials were averaged for each target contrast. Gray area corresponds to stimulus duration. Colored bars indicate the corresponding integration interval of 150 *ms* for averaging neural responses. Integration intervals were defined by incorporating response latencies due to stimulus contrast. Detailed descriptions of response latency is shown in Figure 8

After data preprocessing, imaging data were analyzed to reflect stimulus-evoked neural activity in terms of fluorescence ($\Delta F/F$). VSDI analysis was performed according to the outline in ***Table 1*** and is accompanied with an illustrative example in Figure 7. Example time courses correspond to a cortical site (Figure 7A) in response to a 100 *ms* stimulus that was repeatedly presented four times throughout a second. In the first stage of the analysis, the time course of VSD responses at each pixel was normalized by the mean response from the first 100 *ms* (Figure 7B). VSD responses at the beginning of recording usually provided the maximum response as fluorescence decay in time. Time courses in Figure 7B illustrate 10 blank trials (gray) and 10 trials from a 14% RMS contrast target on a uniform blank background (red). All trials show a systematic decay that is attributed to bleaching effects of VSD molecules. In addition to the systematic decay, individual time

courses from the blank condition show subtle oscillations that increase trial-to-trial variability over time. The increase in variability is presumably due to hemodynamics, respiration, and camera motion that can vary from trial to trial. This is also the case in stimulus conditions (red) but is obscured with stimulus-evoked activity. To minimize trial-to-trial variation from hemodynamics and respiration, each trial was initiated with the onset of the heartbeat signal that was separately monitored form an EKG device. Synchronizing the trial with the onset of a heartbeat effectively reduced the trial-to-trial variability from hemodynamics, as the rate of heartbeats was stable on average. In this example, the stimulus was presented four times, indicated by the four gray regions in the background (Figure 7B). Stimulus-evoked activity is represented by the four bumps in the time course that matches the time of stimulus onset with latency around 40 *ms*, similar to previous studies (Chen et al., 2006; Chen et al., 2012).

The average of blank trial responses are subtracted from all trials to obtain stimulus-evoked responses as shown in Figure 7C. The resulting time course shows four cycles of neural activity that is associated to the stimulus that was presented four times (gray areas). Afterwards, for each trial, the time course of VSDI responses was anchored by the pre-stimulus activity (100 *ms* prior to stimulus onset). Although the stimulus was presented multiple times, I only used the first presentation cycle as the baseline activity after each presentation cycle was not stable . The first presentation cycle was used for all subsequent analyses. The corresponding time course from the first stimulus cycle is shown in Figure 7D.

In Figure 7D, time courses of blank-subtracted and anchored VSDI responses are shown for different target contrasts on a uniform blank background. Each time course is color- coded for different target contrasts, and the corresponding integration time interval

is shown below the time courses. The response latency varied by stimulus contrast and the integration interval was shifted to the initial point of the rising edge (trough).

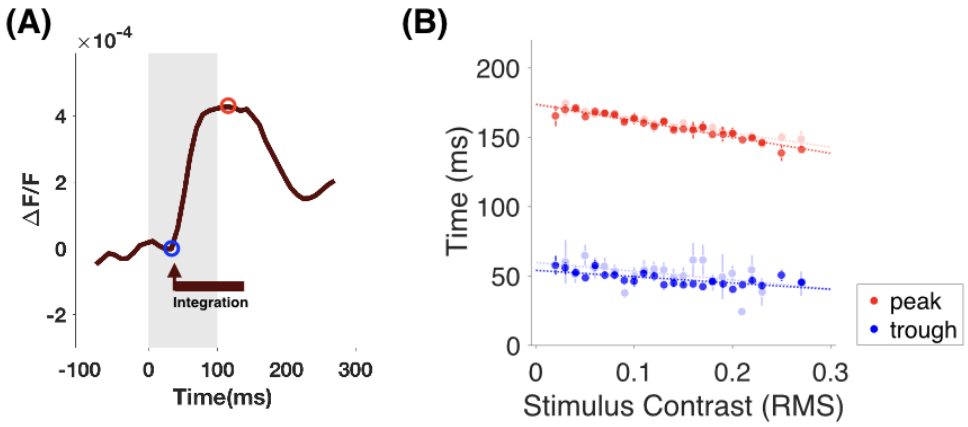

Figure 8: **Integration interval window.** For each trial, the integration time interval shifted by the response latency according to the delay in the rising edge (trough of rising edge). **(A) Peak and trough of the rising edge**: The time course of the strongest target contrast stimulus (14% RMS Gabor) from Figure 7 (monkey B, 11/26/2014) is shown as an example to illustrate how the peak and trough times were estimated. The translucent gray region corresponds to the stimulus presentation. The beginning of the rising edge (trough) was estimated by selecting the point in time where the response amplitude was lowest after stimulus onset. The peak time was selected where the response amplitude was maximum. The integration interval was placed at the trough since it corresponded to the point where the rising edge started. This process was repeated for each trial. **(B) Peak and trough time as a function of stimulus contrast:** For each trial, the stimulus contrast was calculated by the RMS contrast under the target region (~1o diameter). Although background image contrasts were adjusted to four particular values, contrasts under the target region had larger variability and provided multiple data points. Peaks and troughs were estimated from each trial, however, estimations were not successful in all trials. Roughly 50% of all trials provided peak and trough estimations from both monkeys. For monkey A, 463 trials are plotted in faded points. For monkey B, 476 trials are plotted in darker data points. Overall, latencies in both peaks and troughs were nearly the same for both monkeys across stimulus contrast. The response latency was defined by the trough, corresponding to the initial point of the rising edge. The integration interval of 150 *ms* started at the trough.

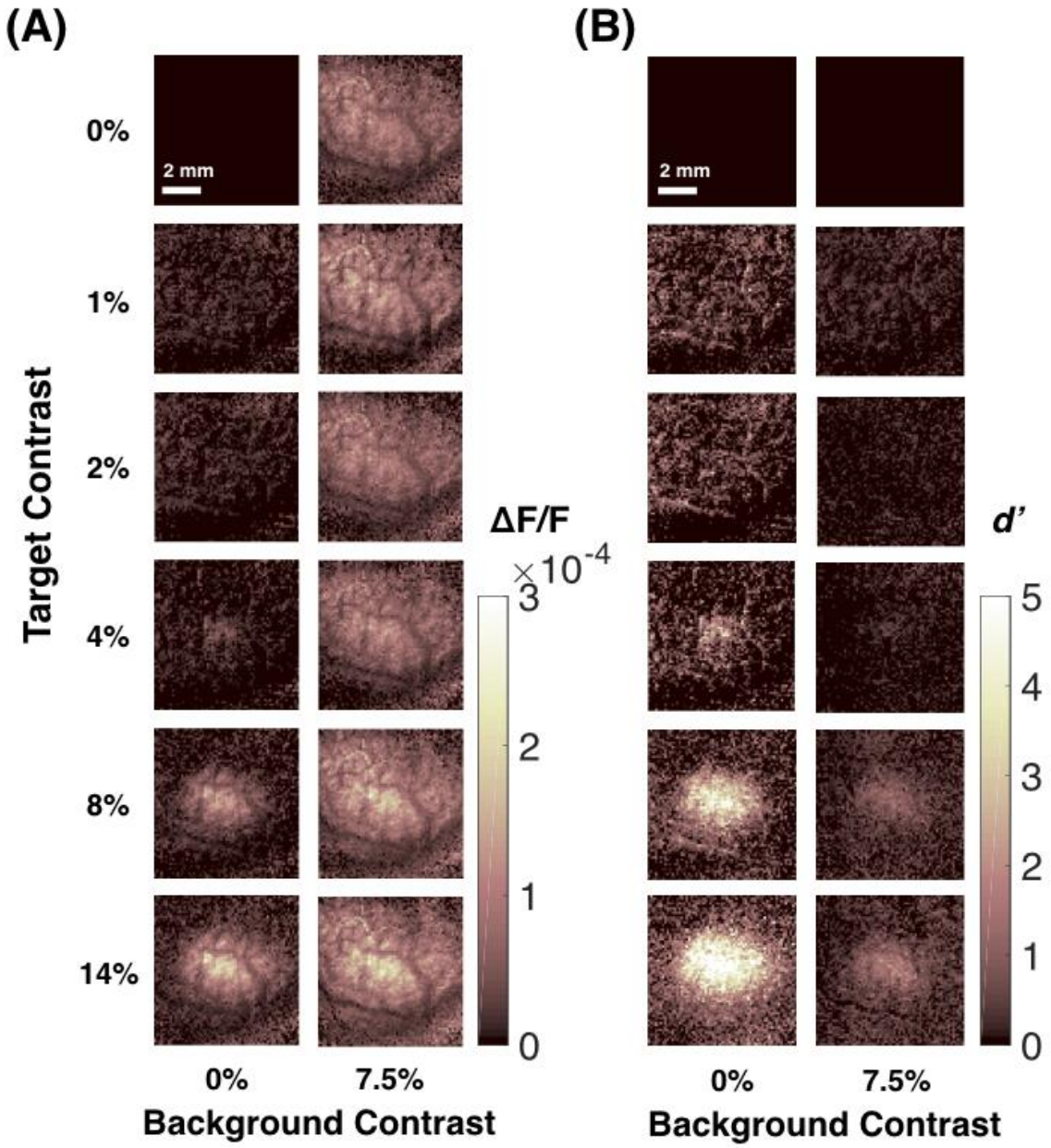

Figure 9: **Example retinotopic responses. (A) Amplitude response maps**: VSDI responses are analyzed to illustrate target-evoked responses across a series of target contrasts (rows) at two different background contrasts (column). **(B) *d'* response maps:** The signal-to-noise ratio (*d'*) of each cortical site was calculated between target-present and target-absent conditions.

The rising edge was estimated by the peak and trough of response time courses in order to approximate response latencies. For each trial, the response time course was characterized by the center 1 mm^2 region of the retinotopic response (Figure 7A). Afterwards, the time course from an individual trial was smoothed by a 30 *ms* moving-window to suppress high-frequency noise when estimating the peak and trough. The trough of the rising edge was estimated by selecting the point in time corresponding to the lowest amplitude from the response sink (dip) right after stimulus onset (Figure 8A, blue circle). The peak time of the rising edge was estimated by selecting the point in time that corresponded to the highest response amplitude before stimulus offset (Figure 8A, red circle). Latencies in peak responses showed a trend with stimulus contrast (Figure 8B, red) that is similar to a previous study but to a lesser extent (Sit et al., 2009). The trend in trough latencies were relatively subtle compared to peak latencies, but indicated a systematic change with stimulus contrast (Figure 8B, blue). Since the trough indicated the point in time where populations were first responsive to the stimulus, the integration interval started where at the beginning of the rising edge—which corresponded to the trough. Therefore, the response latency for each trial was defined by the latency where the trough was observed in the time course (peak latencies did not play a role here). On average, the response latency was around 50 *ms*.

For all conditions, a duration of 150 *ms* was used for the integration interval for the following reasons. First, a fixed interval was used to compare sensitivities across conditions in a fair manner. Although a fixed integration interval might not be the case in the actual decision making process, the goal in this study was to compare neural sensitivities based on the same amount of evidence across all conditions. Second, the reaction time was considered in selecting the duration. From psychophysical measurements in macaques (chapter 2), the median reaction time across all stimulus contrasts was around

200 *ms*. Assuming that responses started to rise after 50 *ms* (Figure 8) and motor preparation for reporting is ~50 *ms*, the temporal interval for response integration window was initially explored with 100 *ms*. In this case, lower contrast responses were nearly indistinguishable with blank conditions. However, since monkeys were capable of detecting low contrast targets on a blank background, the 100 *ms* interval appeared to be underestimating V1 population responses of the low-contrast target. Moreover, reaction times were significantly longer than the median reaction time for lower contrasts, as shown in Figure 3 (> 300 *ms*), suggesting a longer integration interval. Subsequently, the integration interval was increased from 100 *ms* to 150 *ms* for this reason. Using a temporal integration interval of 150 *ms*, VSD responses were integrated at each cortical site, or pixel location, which produced a response map as shown in Figure 9A. The distribution of cortical activity that is sensitive to the target stimulus is better illustrated by taking the signal-to-noise ratio at each cortical site (Figure 9B).

PROPERTIES OF RETINOTOPIC RESPONSES

Characteristics of retinotopic responses are examined in this section to explore how population responses can be read out to identify presence of a target pattern. First, population responses to a small target activate a spatial spread across V1. On average, a 0.84o Gabor target at 1.5° eccentricity evoked a region that was well described by an anisotropic 2D Gaussian with σminor = 1.0 ± 0.06 mm and σmajor = 1.5 ± 0.2 mm (example in Figure 10A). The spatial extent of responses could contribute to inform target-presence for downstream areas when integrating response amplitudes across a summation field (Chen et al., 2006; Sceniak et al., 1999; Sato et al., 2014).

Second, the trial-to-trial variability of retinotopic responses were examined. Mean responses within a region-of-interest (Figure 7A, red square) were sorted across target

contrasts. Mean responses across trials increased with the target contrast while the trial-to-trial variance was stable across target contrasts (Figure 10B). In particular, the trial-to-trial variability from target-present trials and noise trials (blank screen) was not significantly different. This was also documented in Chen et al. (2006) where the authors suggested that population dynamics between the mean and variance reveal different characteristics in comparison to individual neurons where the mean and variance of spike counts are proportional to each other. From this study and Chen et al. (2006), VSD recordings indicate that the relationship between the mean and variance can be significantly weaker in comparison to single unit recordings.

In addition to the response correlation across conditions, the response correlations were also spatially widespread across stimulus-evoked regions in V1. The average stimulus-evoked response was subtracted for each condition and the residuals from each trial was used to measure spatial correlations. When spatial correlations between pairs of cortical sites were measured as a function of their spatial separation, the average pairwise correlation was relatively high and decreased with cortical distance (Figure 10C). Across trials, the spatial correlation structure was also not significantly different between target-present trials and target-absent trials. The widespread spatial correlations were described by an additive Gaussian noise that was stimulus-independent and spatially correlated with an exponential decrease from the center of the imaging window. A potential explanation for the high stimulus-independent correlation could be attributed to the nature of increasing correlation in a large network of neurons. Although pairwise correlations at the level of individual neurons are weak, the aggregate correlation across the population could be substantially higher (Chen et al., 2006; Zohary et al., 1994).

In summary, three major properties of retinotopic responses were observed when presenting a small target pattern in a uniform blank background. First, stimulus-evoked

activity was spatially distributed across a local area in the V1 retinotopic map and the peak response increased with the stimulus contrast. Second, the trial-to-trial variability of population recordings was constant across target-present trials and target-absent trials. This could potentially indicate that the strength of the target signal can be better represented in mean responses across trials since the trial-to-trial variance is relatively constant. Lastly, although the overall activity increased with stimulus contrast, spatial correlations between cortical sites were independent of the stimulus.

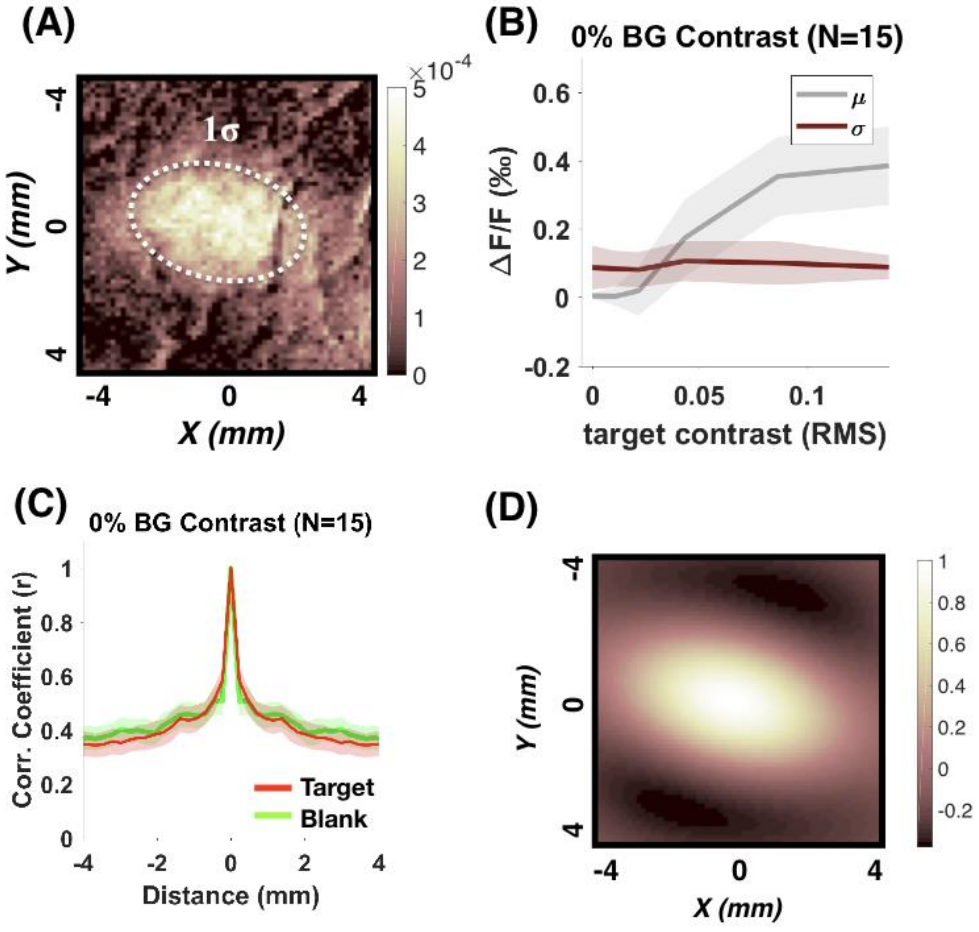

Figure 10: **VSDI response properties from a target on a uniform background**. **(A) Target-evoked regions**: a 0.84^o Gabor target at 1.5^o eccentricity evoked a region that was well described by an anisotropic 2D Gaussian with σ_{minor} = 1.0 ± 0.06 *mm* and σ_{major} = 1.5 ± 0.2 *mm*. **(B) Trial-to-trial variability**: mean responses within a region-of-interest (Figure 2A, red square) were sorted across target contrasts. **(C) Spatial correlations**: correlations between pairs of cortical sites were measured as a function of their cortical distance (averaged across 15 experiments from monkey B). **(D) Spatial weights for optimal pooling:** spatial weights were obtained by spectral whitening. A decision variable for each trial was derived by applying this map of weights to a trial's retinotopic map. Distributions of decision variables are then used to calculate *d'* values to derive a neurometric function (Figure 15).

In the case of target patterns on natural images backgrounds, similar response properties were observed but with minor differences. First, trial-to-trial variability of population responses were constant and were similar to the variability in a uniform background (Figure 11). However, the mean response across different target contrast levels were relatively less distinguishable in comparison to the case of a uniform background. As the target is added to the background, the combined stimulus contrast appeared to push retinotopic responses to a range where population responses saturates (Figure 11A). Second, widespread spatial correlations were observed as well, but as background contrast increased, the structure of spatial correlations appeared to deviate between target-absent and target-present conditions (Figure 11B). When a high-contrast target was added to natural image backgrounds, spatial correlations across cortical sites appeared to be relatively lower than the case when the same backgrounds were presented without the target. At the current stage of this study, the significance of background effects in spatial correlations is not completely clear. However, since a major difference in response properties was not noticed, the same pooling approach was applied to both uniform background and natural image background stimuli as an exploratory attempt.

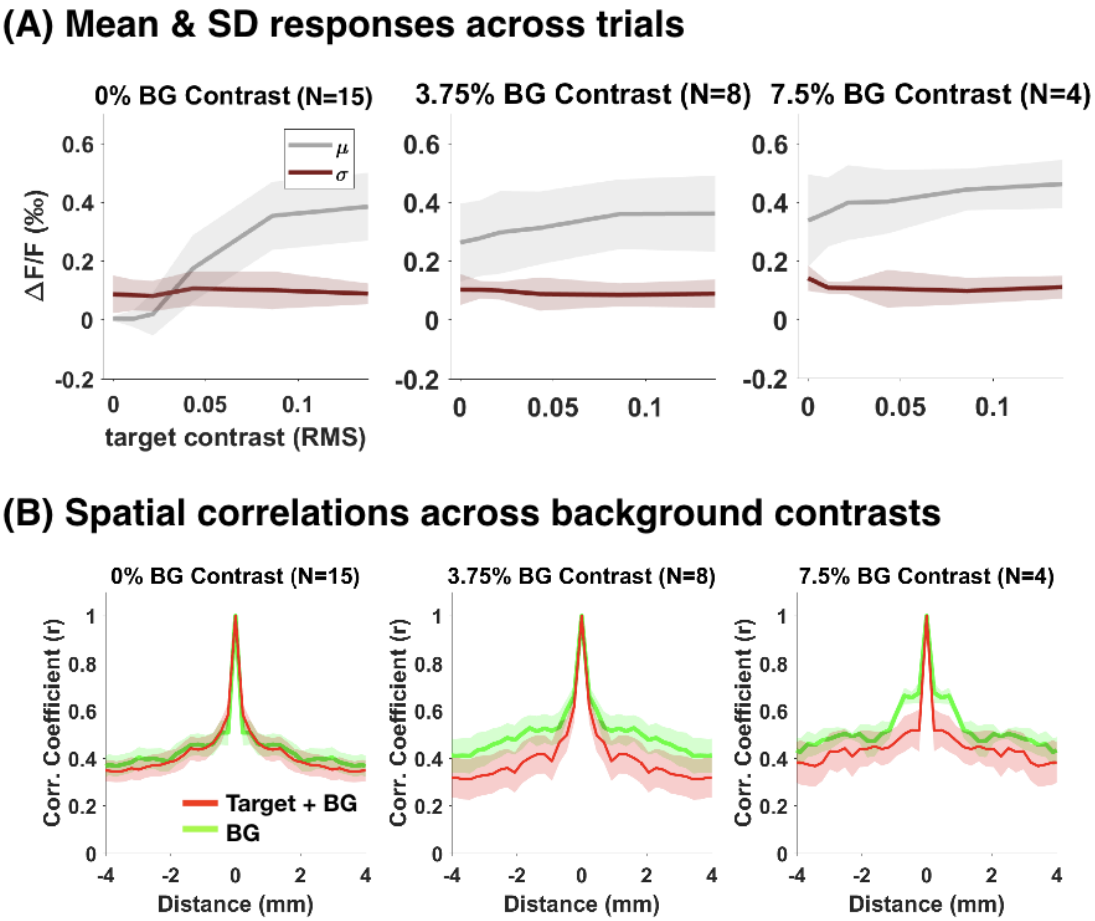

Figure 11: **VSDI response properties from a target on natural image backgrounds**. Trial-to-trial correlations across background contrasts and cortical distance from the center are shown. **(A) Mean and SD responses across background contrasts:** The same analysis from Figure 10B was applied to conditions across background contrasts. Trial-to-trial variability are nearly identical across background contrasts, whereas mean responses approached saturation with increasing background contrast (same imaging experiment with previous figures; monkey B, 11/26/2014). **(B) Spatial correlations**: Noise correlations between pairs of cortical sites were measured as a function of their cortical distance, as in Figure 10C, across different background contrasts. For each experiment, spatial correlations are calculated from background-only conditions and conditions where the highest-target contrast was added to natural backgrounds. Both conditions had 10 trials. Shaded area indicate one standard error.

NEURAL SENSITIVITY OF RETINOTOPIC RESPONSES

Based on statistical properties of VSD measurements, an optimal set of weights were developed by Chen et al. (2006) that maximizes target discriminability in a uniform background. VSDI recordings revealed that spatial correlations at the retinotopic scale were independent to the stimulus and were well described by Gaussian noise. Under this noise model, spectral whitening was applied to maximize the discriminability between target-present and target-absent responses on a uniform blank background. This approach is optimal under certain circumstances: (a) responses at each cortical site is a Gaussian random variable, (b) responses at each site are independent across trials, and (c) the covariance across cortical sites do not depend on the stimulus. For responses in a uniform background, the optimal set of weights across the cortical space resulted in an antagonistic center-surround arrangement that effectively enhanced signals concentrated in the target-evoked region while suppressing surrounding activities (Figure 10D).

In the case of background conditions, weights were also derived through spectral whitening. However, optimality is not guaranteed since previous assumptions of response characteristics from a uniform background may not hold in the case with background stimuli. Figure 11B suggests that spatial correlations between target-absent and target-present trials might not be entirely independent. In such case, backgrounds might affect the covariance across cortical sites and the correlated variability in population responses is not completely independent of the stimulus. Therefore, these weights are likely to be suboptimal in the presence of backgrounds and will be referred to retinotopic weights where these spatial pooling weights are applied to the response map of each trial to derive a decision variable. Particularly, the inner product between the antagonistic arrangement of spatial weights and the two-dimensional response map was used to pool a scalar decision variable to represent the target signal. A distribution of decision variables was obtained by

repeating trials at a particular trial contrast. Retinotopic weights were separately derived for each experiment and was cross-validated by holding out a trial of interest and independently deriving the weights from all other trials.

At each target contrast, the signal-to-noise ratio (*d'*) was calculated from decision variables between trial-present and trial-absent trials. Across multiple target contrasts, a series of d'-values were obtained, and the Naka-Rushton function (equation 6) was fitted to describe d'-values as a function of target-contrast. By convention, the Naka-Rushton function was used to describe the mean response as a function of contrast. A neurometric function was derived by converting *d'* values to percentages of correct trials (Figure 15, blue). This approximately approaches an ideal observer's performance as the optimal criterion is defined by *d'* values (assuming equal-variance Gaussian distributions of decision variables). The neural threshold was defined to be the target contrast at d'=1 following the same convention with the behavioral experiment. The first column of plots in Figure 15 show examples of neurometric curves superimposed to psychometric curves.

$$d'(c) = d'_{max}\frac{c^n}{c^n + c_{50}^n} \tag{6}$$

PROPERTIES OF ORIENTATION COLUMN RESPONSES

Detection sensitivities of V1 population responses are also explored at the scale of orientation columns. Most V1 cells are selective to orientation and their columnar organization is a hallmark feature of V1 in primates. Although the functional role of orientation columns is unclear (Horton and Adams, 2005), neural signals conveyed in these columns could be relevant in identifying a target pattern from its background. To evaluate this idea, detection sensitivities to a target in natural backgrounds are estimated for population responses at the scale of V1 orientation columns.

To extract neural activity at the scale of orientation columns, a separate imaging experiment was conducted to isolate population signals that were sensitive to orientation. In this separate experiment, VSDI response maps were collected from trials in which high contrast sine-wave gratings were presented at various orientations to determine population signals that were selective to orientation.

Two monkeys (monkeys B and A) were trained to maintain fixation while static orientation gratings were briefly flashed at 5 Hz. The Fourier amplitude that matched the stimulus presentation frequency (5 Hz) was computed at each pixel location to derive a VSDI response map for each trial. Pulsed stimulus presentation was shown to enhance the signal-to-noise ratio and was used here to achieve better estimates of orientation column activity (Chen et al., 2012).

I first looked into VSDI response maps from grating stimuli and compared this to a control condition (blank trials). The amplitude spectrum of VSDI recordings revealed a distinct spatial profile from orientation gratings when compared to blank trials (Figure 12A). The spectral analysis was conducted in the following manner. First, average responses from blank trials were first subtracted from all conditions (blank subtraction). Next, responses from each orientation was averaged and a Hamming window was applied to minimize ringing effects. Afterwards, the two-dimensional amplitude spectrum was computed. Lastly, the two-dimensional spectrum was averaged radially to derive the one-dimensional amplitude spectrum.

Based on the one-dimensional amplitude spectrum, populations responding to orientation gratings showed increased activity in the frequency range of 0.8 - 3 *cyc/mm*, which corresponds to a periodicity of 0.33 - 1.25 *mm* at the surface of the cortex. Neural activity concentrated in this frequency-band was isolated by subtracting the baseline 1/f trend (Figure 12B, *blue*) and fitting a bandpass filter (Figure 12B, *red*) according to

equation (7). In equation (7), α is the peak amplitude, β is the slope of the falling edge, f_0 is the frequency at the peak amplitude, f_{50} is the half-peak bandwidth, and c is a constant to fit the offset. The center frequency and bandwidth of orientation specific activity was robust across both monkeys. The center frequency of 1.3 *cyc/mm* and bandwidth of 1.1 *cyc/mm* is in agreement to previous findings using the same approach (Chen, Palmer, and Seidemann, 2012).

$$f(x) = \frac{\alpha}{1 + (\frac{|x - f_0|}{f_{50}})^{\beta}} + c \tag{7}$$

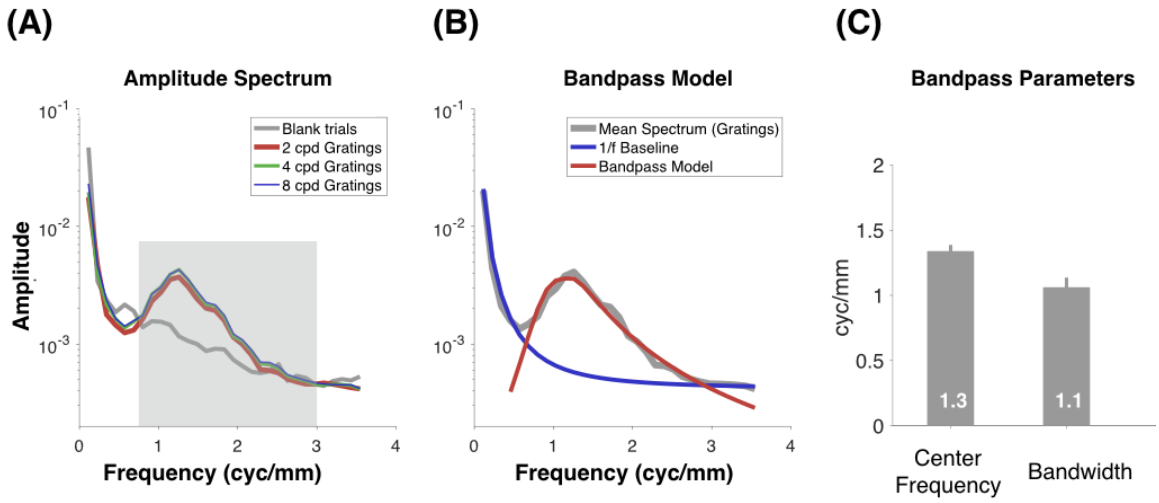

Figure 12: **One dimensional amplitude spectrum of orientation column responses**. **(A)** The 1D amplitude spectrum from a single experiment is shown for orientation gratings and blank conditions. Neural responses to orientation gratings showed increased amplitudes that were concentrated in the frequency range of 0.8 - 3 cyc/mm in the cortex, which corresponds to a pattern of neural activity having a periodicity of 0.33 - 1.25 mm. **(B)** Neural activity at this frequency band was isolated by subtracting the 1/f structure (blue) and fitting a bandpass filter (red) according to equation (7). **(C)** Averages of the center frequency and bandwidth from monkey B (7 experiments) and monkey A (5 experiments).

To further validate whether neural activity at this band is specific to orientation columns, pair-wise correlations between bandpass-filtered responses were computed (Figure 13). To verify whether band-passed responses reflect orientation columns, the correlation between pairs of 2D response maps were computed as a function of their stimulus orientation difference. If band-passed responses do not reflect orientation column activity, response maps would be independent of differences in stimulus orientation and the correlation coefficient between pairs of maps would be close to zero. I find that this is not the case as response maps were correlated for similar stimulus orientations (when orientation differences are near zero), and anti-correlated when the stimulus orientation at 90º (Figure 13B). This provides evidence that band-passed (0.8 – 3 *cyc/mm*) response maps reflect neural activity at the scale of orientation columns.

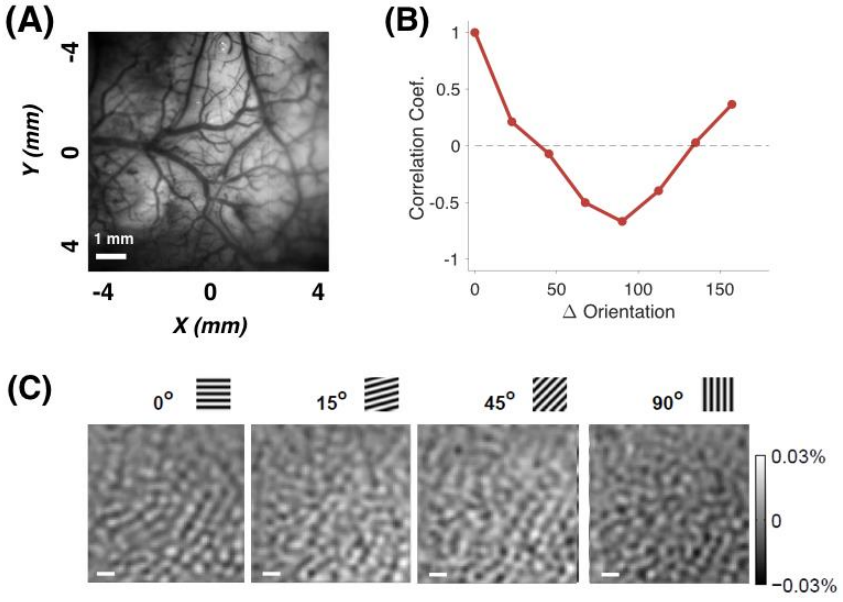

Figure 13: Pair-wise correlations of columnar responses across stimulus orientation. **(A)** Region of interest (ROI) for VSD imaging. Retinotopic responses are first analyzed from this ROI and band-passed to isolate orientation column maps. **(B)** Correlations between pairs of band-passed maps as a function of the difference between a pair of stimulus orientation. **(C)** Four examples of orientation column maps that were used in (B).

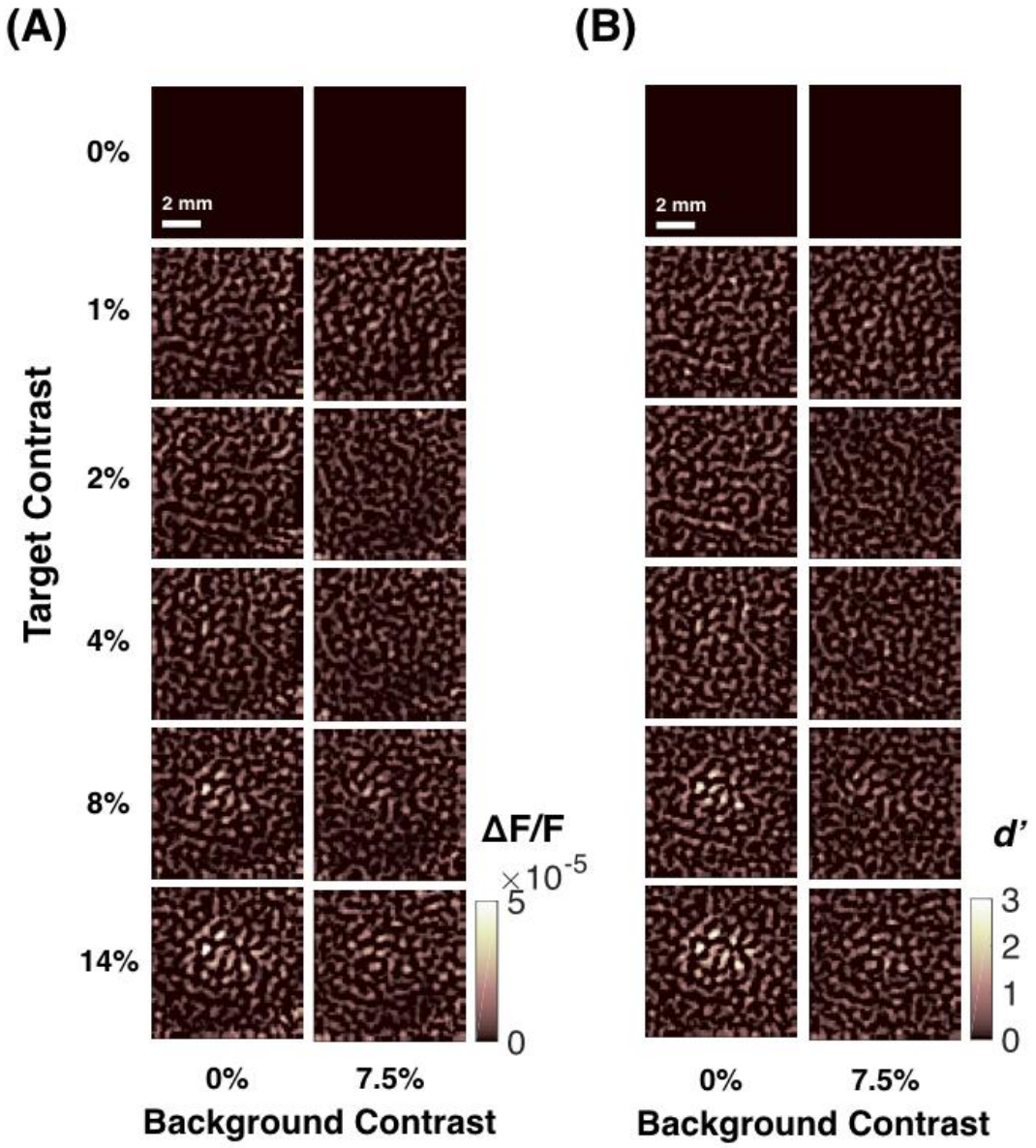

Figure 14: Example response maps of orientation columns. Orientation column responses were extracted by band-passing (Figure 12) retinotopic responses from Figure 9. **(A) Columnar response maps**: VSDI responses to orientation columns are arranged across a series of target contrasts (RMS) at two different background contrasts (uniform background and 7.5% contrast background). Brighter values represent higher amplitudes of columnar responses. **(B) *d'* response maps**: The signal-to-noise ratio (*d'*) of each cortical site was calculated between target-present and target-absent conditions from (A).

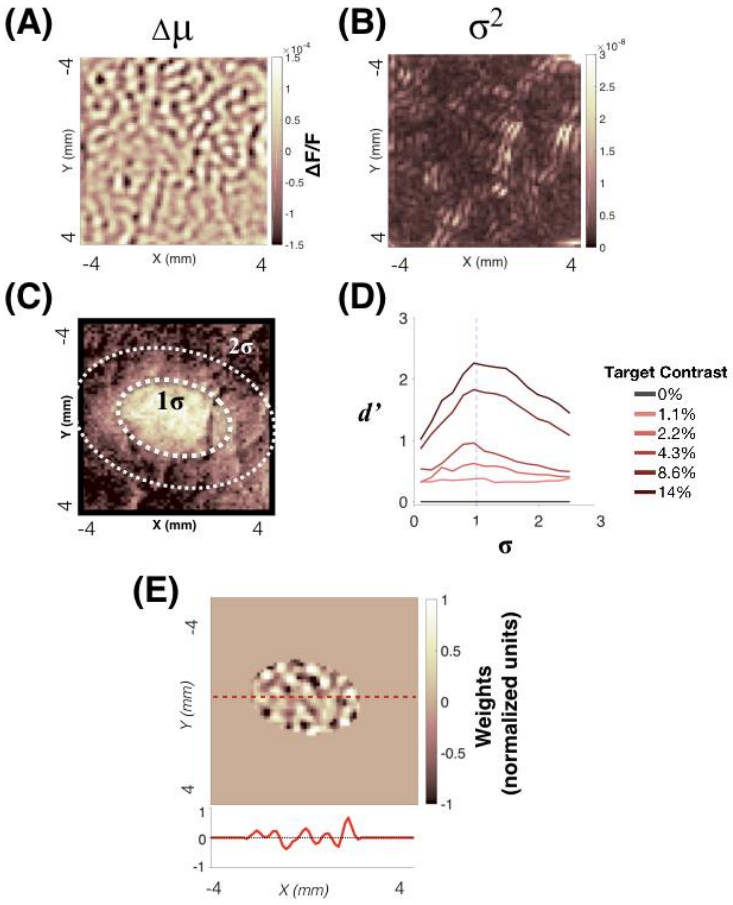

Figure 15: **Orientation column template**. For each experiment, orientation column templates were derived from a separate block using orientation gratings. The template is essentially a reliability map that applies more weight to cortical regions at the scale of orientation columns and target-evoked regions. The following components are used to derive a template. **(A)** Mean differences are first computed by subtracting responses between two orthogonal stimuli (horizontal vs vertical grating). **(B)** Each location in the mean difference map from (A) is normalized by its reliability. **(C)** The reliability-weighted response map is now masked by the target-evoked region. The retinotopic spread of the target stimulus was fitted by a 2D Gaussian and the spatial extent of recruiting columns was limited to the first standard deviation of the Gaussian ellipse. **(D)** Effect of the retinotopic extent: *d'* values (between target-absent and target-present trials) are shown as a function of the spatial extent of columnar responses in units of standard deviations. The spatial extent of the first standard deviation was used for the template. **(E)** Orientation column template: The resulting template is consisted of spatial weights that correspond to orientation columns that prefer the target orientation (horizontal in this example).

NEURAL SENSITIVITY OF ORIENTATION COLUMN RESPONSES

A matched-template model is used to evaluate the detectability of the target stimulus in orientation column responses. The matched-template model is optimal when the target is known and masking noise is additive and Gaussian (in general, independent and identically distributed) (Peterson et al., 1954; Green and Swets, 1966). In the case of columnar responses, the statistical structure of noise is not well-known and optimal performance is therefore not guaranteed. However, the matched-template model is a principled model that is linearly tractable in understanding the data. In addition, alternative candidate templates could be used to interpret how different approaches in pooling columnar responses affect target-detectability. Nonetheless, the template is considered to be a prototype of orientation column responses that convey orientation features of the target stimulus. Subsequently, the orientation column template in this study is constructed from V1 orientation columns that prefer a horizontal orientation since the target stimulus is a horizontal Gabor.

For every experiment, a separate imaging session was conducted to derive the template. The orientation column template is constructed as a 2D map of spatial weights that selectively recruit columns with preferred orientations that matched the target orientation. This was implemented by recording V1 responses while presenting two orthogonal orientation gratings (0^o versus 90^o). Imaging data was first centered and the bandpass filter was applied to extract columnar responses. Orientation columnar responses from the two orthogonal stimuli were subtracted from each other (Figure 15A) and weighed by the reliability, or normalized by the variance, at each cortical location (Figure 15B). Afterwards, the spatial spread of the target-elicited region in the cortex was considered in restricting the size of the template. The retinotopic spread of the target stimulus was fitted by a 2D Gaussian and the spatial extent of recruiting columns was limited to the first

standard deviation of the Gaussian ellipse (Figure 15C). The spatial extent of the first standard deviation was used because the discriminability between target-present and target-absent responses was maximal at this spatial extent across a range of target contrasts (Figure 15D). The resulting template consisted of spatial weights that prefer horizontal orientation columns and are spatially restricted by the first standard deviation of the retinotopic spread of the target stimulus (Figure 15E).

On each trial the matched-template model applies the template to the band-passed trial response to derive a decision variable. The decision variable (*DV*) is computed by the dot product of the template, *T(x,y)*, and the columnar response of a trial, *I(x,y)*:

$$DV = \sum_{x,y} T(x,y) \cdot I(x,y) \tag{3}$$

At each target contrast the signal-to-noise ratio (*d'*) was calculated between decision variables from trial-present and trial-absent trials. Similar to the retinotopic decoder, a series of *d'* values were obtained across multiple target contrasts and the Naka-Rushton function (equation 1) was fitted to describe *d'* values as a function of target-contrast. The neural threshold was defined to be the target contrast at *d'*=1 following the same convention with the behavioral experiment and the retinotopic decoder. Neurometric functions were derived by converting d' values to percentages of correct trials (optimal criterion = *d'*/2 for an equal-variance Gaussian model). Examples of neurometric functions for orientation column responses are shown in Figure 16 (*red*).

In summary, V1 population responses at the scale of the retinotopic map and orientation columns were recorded using VSDI in a fixating monkey that was not trained to perform a detection task. Neural thresholds were estimated in separate ways for the retinotopic response and orientation column responses. At the scale of retinotopic

responses, a set of spatial weights was derived to whiten spatial correlations in VSDI measurements. For orientation column responses, a matched-template model was used to linearly combine responses from orientation columns that preferred the target orientation. Neurometric functions from retinotopic and columnar responses were estimated to quantify target detectability, or a neural detection threshold. This quantity is compared to behavioral detection thresholds measured from separately trained monkeys to test whether detection performances in natural backgrounds can be predicted from V1 population responses.

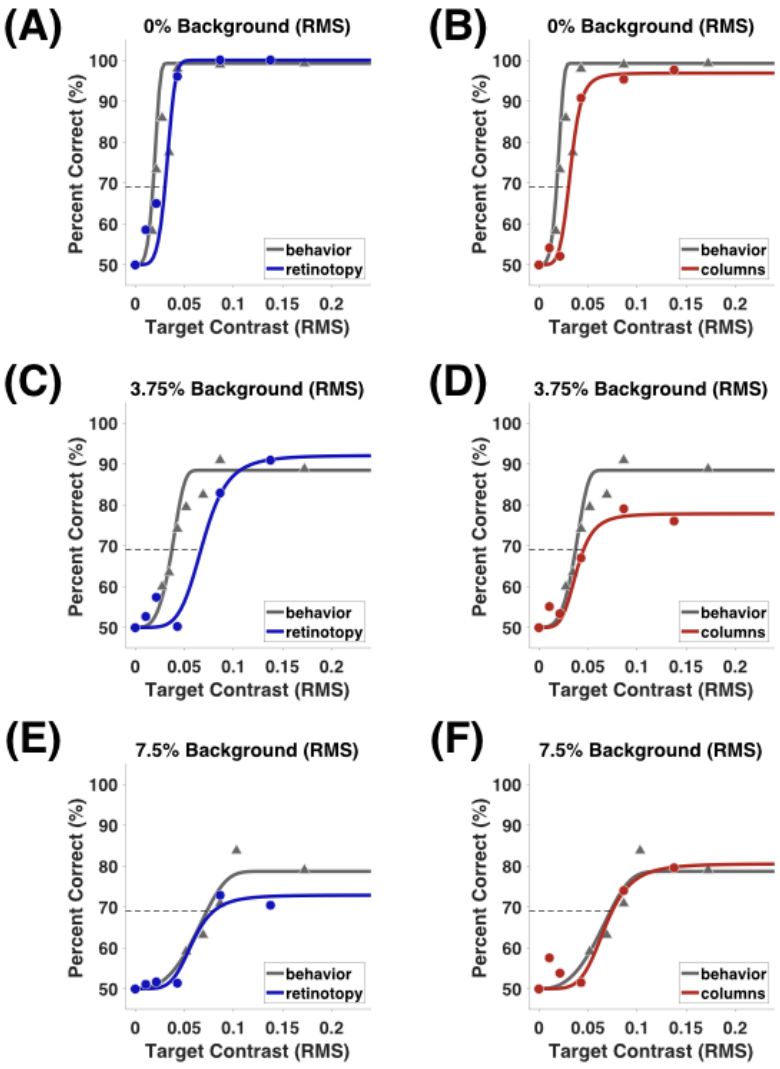

Figure 16: **Psychometric versus neurometric functions**. Representative examples of neurometric functions from retinotopic (*blue*) and columnar responses (*red*) are superimposed on psychometric functions (chapter 2) across three background contrasts (rows). Retinotopic and columnar performances are listed in columns. Contrast detection threshold is defined at 69% correct (*d'* = 1) for both psychometric and neurometric functions. Psychometric and neurometric functions were measured from separate monkeys.

RESULTS

Figure 17 shows the overall trend of detection sensitivities in retinotopic responses, orientation column responses, and behavior. Monkey B was experimented with a stimulus that was on for 100 *ms* and off for 150 *ms* (Figure 17A,B), whereas the stimulus for monkey A was on for 200 *ms* and 300 *ms* off (Figure 17C,D). An integration duration of 150 *ms* was used in calculating neural responses for both monkeys. Experiments were included in the analysis only when both retinotopic and columnar responses surpassed threshold levels (*d'* > 1). Each data point in Figure 16 is the grand mean of thresholds and the number of samples are matched for retinotopic and columnar thresholds. Experiments that were included in the analysis is listed in Figure 18 for monkey 1 and Figure 19 for monkey 2.

In terms of retinotopic responses, neural thresholds elevated with background contrast and were always higher than behavioral detection thresholds (gray line). Higher neural thresholds could be attributed to underestimating neural sensitivities when using a suboptimal size of the target stimulus. In a previous study, Chen et al. (2006) demonstrated that VSD responses to a ~2° Gabor target (σ=0.25-0.33°, SF=1.4-1.7 *cpd*) could be pooled to outperform behavioral detection performances. In this thesis, a ~1° Gabor target was used and the corresponding neural thresholds were systematically higher than behavioral detection thresholds. The exact extent of signal attenuation to a smaller target is unclear but can be speculated by comparing the VSD contrast response function (CRF) between

the two target sizes. In Figure 22, VSD response amplitudes from a cortical site (Figure7A) are plotted as a function of target contrast. The RMS contrast was matched for the two target sizes and recorded under the same recording configuration. The semi-saturation (C50) value was significantly higher for the smaller target, suggesting that VSD measurements could be less responsive to smaller stimulus size and effectively increasing neural thresholds.

Aside from the effect of target size, neural sensitivities between subjects showed differences. Neural sensitivities of the second monkey were systematically lower, which is most likely due to deteriorating conditions of the cortical tissue at the time of imaging. Cortical tissue health was affected by extraneous factors which is likely to have impacted the overall imaging quality.[1]

In terms of orientation column responses, detection thresholds were similar to retinotopic responses except for the highest background contrast. For monkey B, orientation column thresholds were lower than retinotopic thresholds. However, only one experiment is reported here and therefore this finding is not conclusive due to the lack of statistical power (Figure 17A, far right data-point). For monkey A, column responses reached threshold levels whereas retinotopic responses did not. This might be consistent to monkey B, where orientation columns are more sensitive to the target stimulus in natural backgrounds. However, this can only be speculated due to deteriorated tissue health of monkey A and lack of statistical power in monkey A as well.

In order to illustrate the relative trend of detection sensitivities, data points were normalized by the first data point, or the threshold from a uniform background (Figure 17B, D). Across two monkeys, the relative trend reveals a better correspondence between

[1]Patching experiments were conducted in the same weeks of imaging

behavioral and retinotopic performances. Relative performances of orientation columns were less susceptible to background contrast masking.

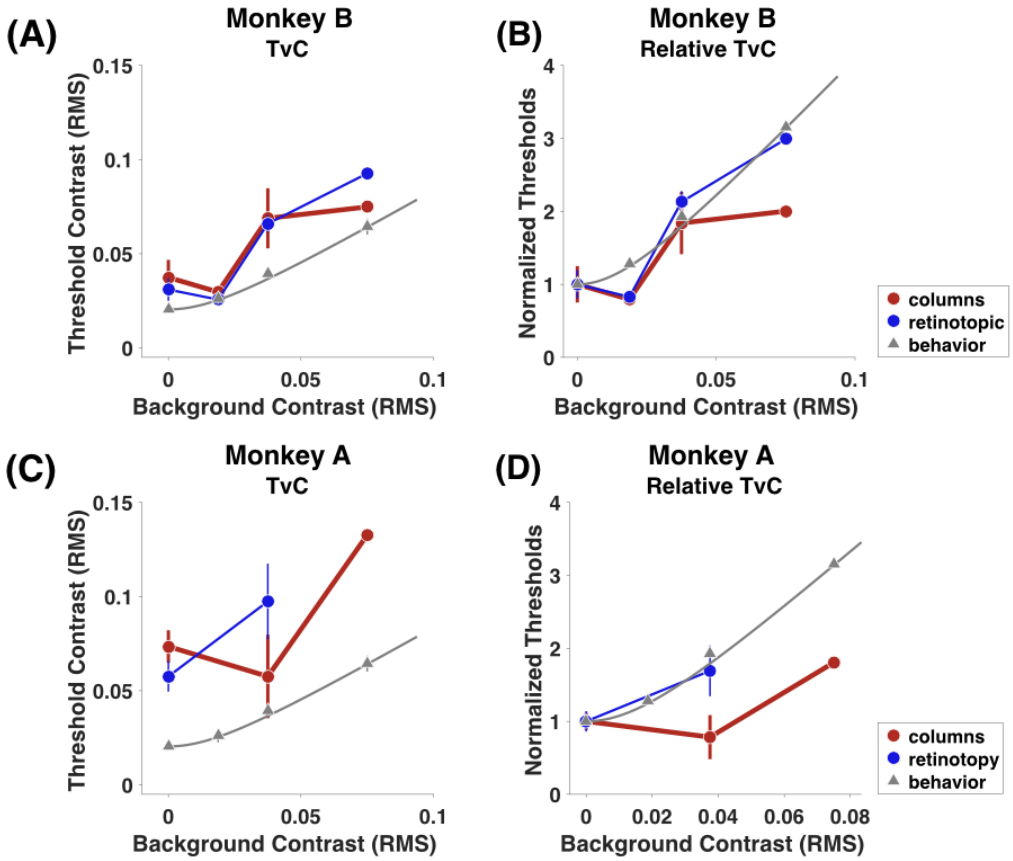

Figure 17: **Summary of detection sensitivities.** Behavioral and neural detection performances are plotted for monkeys B and A. **(A) Thresholds versus background contrast:** Contrast detection thresholds are plotted as a function of background contrast. Each data point is the grand mean across imaging sessions from monkey B. The number of samples for each data point is matched between columnar responses and retinotopic responses. **(B) Relative trend of thresholds:** To better visualize performance trends, each set of threshold values were normalized by the first data point in (A). **(C)** Same format with (A) for monkey A, however, the total number of background contrasts that was experimented in monkey A was three instead of four (monkey B). Also, for monkey A, retinotopic responses did not reach threshold at the highest background contrast whereas columnar responses reached threshold. Resulting in a one less data point. **(D)** Same as (B) for monkey A. (Figures A&B).

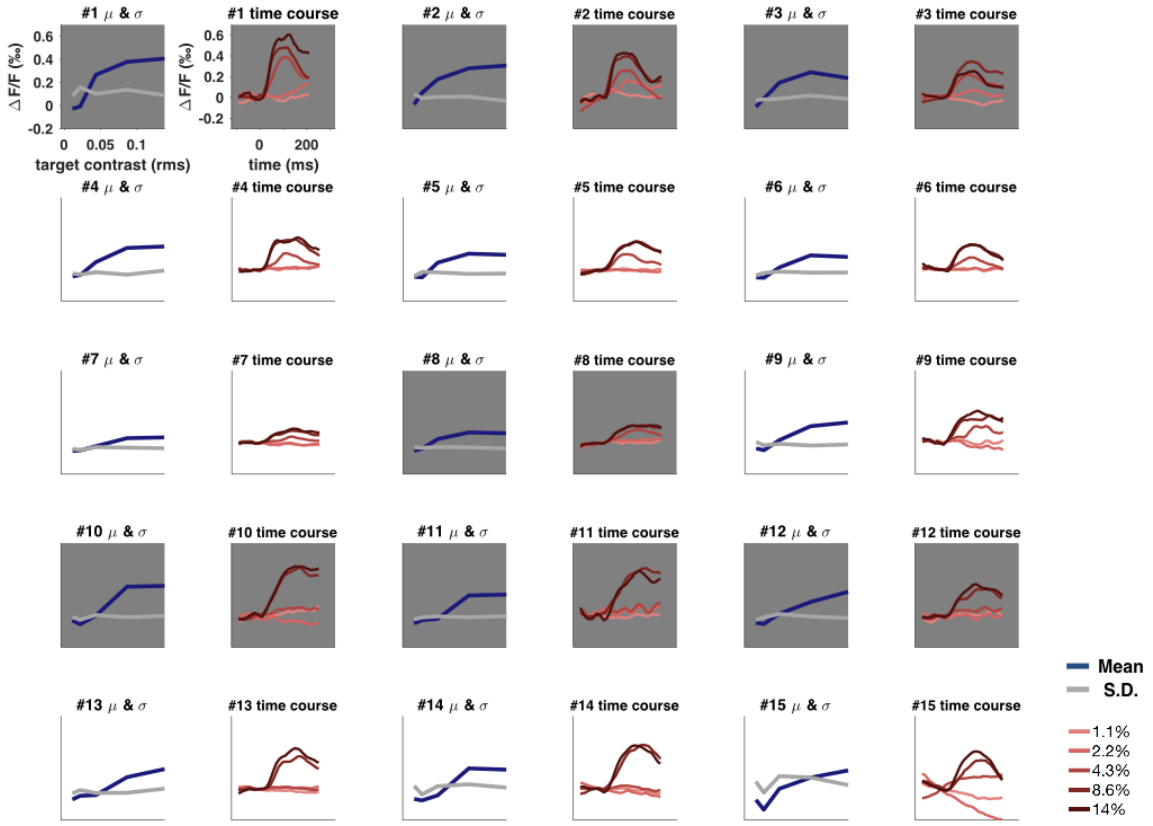

Figure 18: **Retinotopic responses of all experiments from monkey B (supplement for Figure 16)**. A total number of 15 imaging sessions were conducted from monkey 1. Summary of response amplitudes for an individual experiment is plotted in a pair of plots. The first subplot shows the mean and standard deviation of amplitudes across target contrasts on a uniform background. Response amplitudes are calculated from the central 1 x 1 mm2 area of the stimulus-evoked region. The second subplot depicts the time course from the first stimulus presentation cycle across target contrasts. ***Data inclusion criterion***: for each imaging session, uniform background and natural background conditions were collected. The experiment was included in the analysis only when both retinotopic and columnar responses reached threshold in the uniform background. Grey backgrounds indicate sessions that were excluded because either responses at the scale of the retinotopy or orientation columns did not reach threshold.

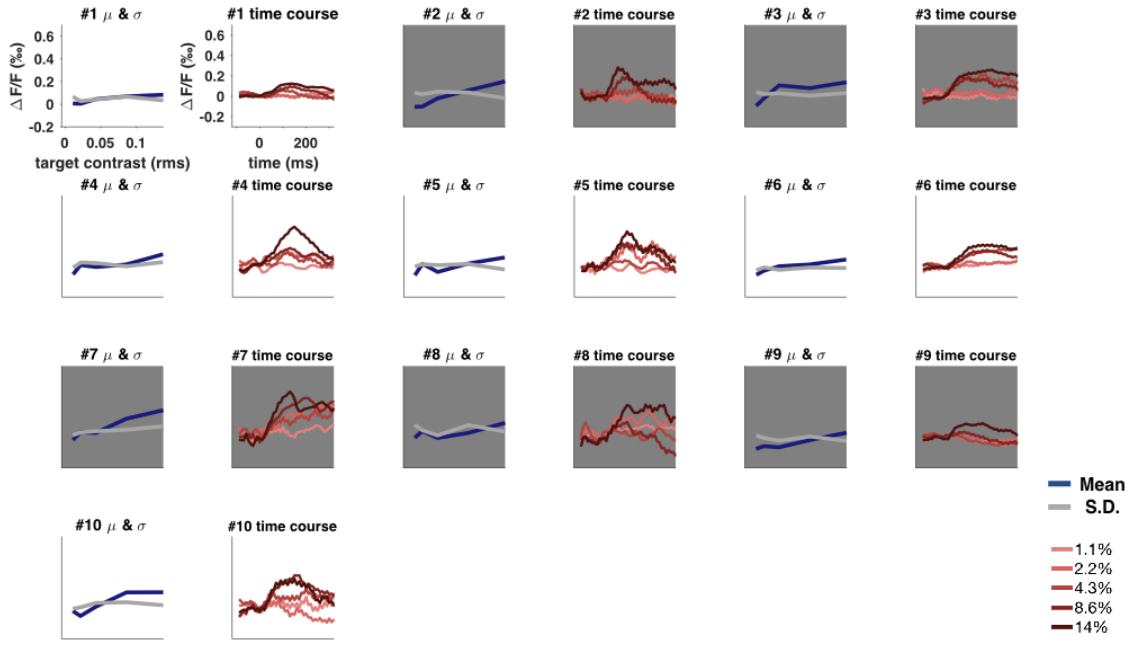

Figure 19: **Retinotopic responses of all experiments from monkey A (supplement for Figure 16)**. A total number of 10 imaging sessions were conducted from monkey 2 and the follows the same plotting format with Figure 18.

RESULTS WITH IMAGE STABILIZATION

As imaging data is collected for an extended period of time (> 1 second), movement in the imaging apparatus (either the monkey or camera) introduces misalignments in imaging frames and degrades the average signal due to blurring effects. To correct for such artifacts, an additional pre-processing procedure was added to correct for motion blurring. Before estimating motion effects, the illumination profile was incorporated to better estimate motion vectors within the imaging area. Afterwards, for each trial, motion effects were estimated across time. A reference frame was chosen at the beginning of the trial and lateral motion vectors were estimated across all frames. Motion effects were corrected based on these estimations and was capable of enhancing within-trial signal qualities.

The overall result with image stabilization is shown in Figure 20. Stabilized results are shown in dotted lines and superimposed to previous results from Figure 16. In effect,

image stabilization elevated thresholds in retinotopic signals but lowered thresholds from columnar signals. Moreover, in the case with orientation columns, the total number of experiments that reach threshold levels increased when applying image stabilization (Figure 20 versus Figure 18). Previously, 7 out of 15 experiments did not reach threshold levels in the case of orientation column responses (Figure 18). Among the 7 experiments that did not reach threshold, 3 experiments were recovered with motion correction. Overall, only 4 out of 15 experiments did not reach threshold with the help of image stabilization, as shown in Figure 20.

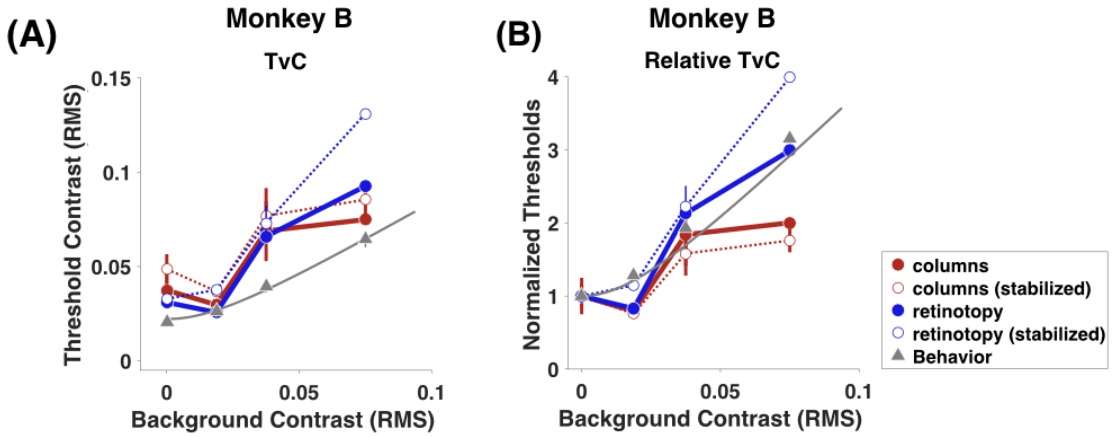

Figure 20: **Performance summary with image stabilization**. An additional pre-processing step was added to correct for motion during an imaging trial. This was explored in monkey B only and the results are shown in dotted lines for (A) threshold versus background contrast, and (B) normalized thresholds as a function of background contrasts with image stabilization. Original results without image stabilization are from Figure 17 (thick lines). Overall, an increased number of imaging sessions could reach threshold levels with image stabilization as shown in Figure 21.

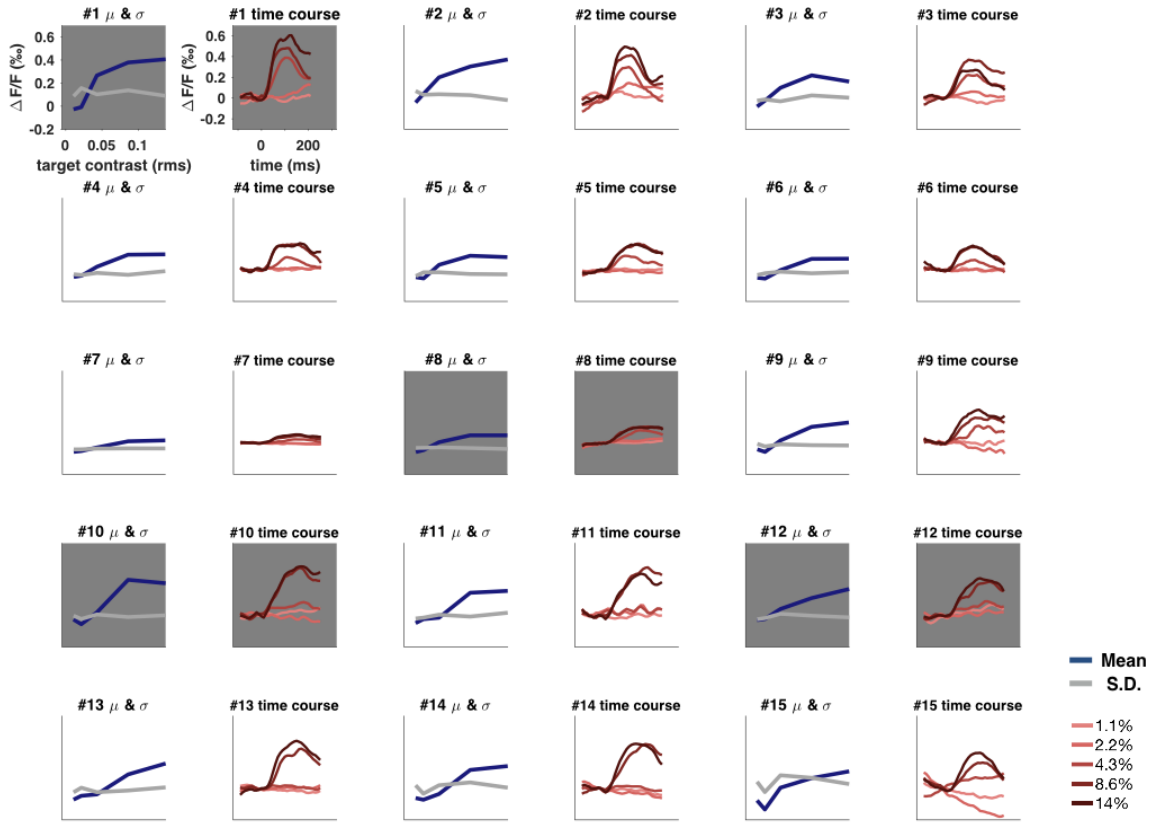

Figure 21: **Image stabilized experiments that were used in Figure 20**. Individual experiments that were used in reporting threshold values in Figure 20 is shown. Figure 18 is repeated with image stabilization. The total number of experiments that satisfied the data inclusion criterion increased with image stabilization.

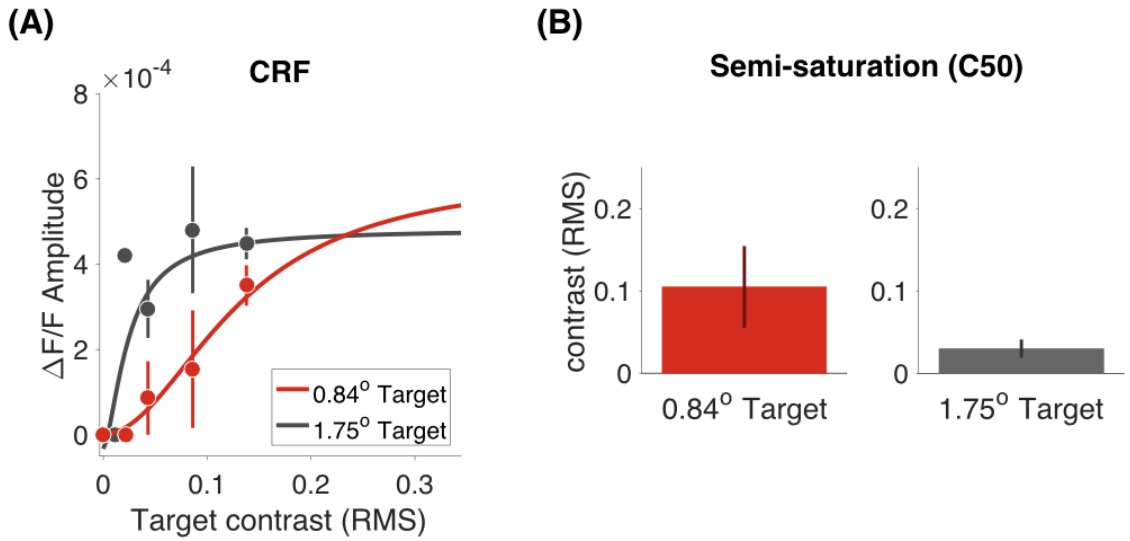

Figure 22: **Target stimulus size effects in VSDI recordings. (A)** Response amplitudes from a cortical site (Figure7A) are plotted as a function of target contrast for two target sizes. The RMS contrast was matched for the two target sizes and recorded under the same recording configuration. **(B)** The semi-saturation (C50) value was significantly higher for the smaller target, suggesting that the sensitivity of VSD measurements could be lowered with a smaller target stimulus size. The discrepancy in the neural threshold values from Chen et al. (2006) could be attributed to this observation.

Chapter 4: Discussion

In this thesis I measured V1 population responses and compared them to behavioral detection performances in natural image backgrounds. Psychophysical data from macaques were first collected and compared to human data in the same experiment to determine whether the macaque is an appropriate animal model in studying human detection behavior. The relative trend in detection performance was similar between humans and macaques, but with lower sensitivities in macaques.

Given the similar detection performances of humans and macaques, neural population activities in macaque V1 were measured to gain insight to underlying neural mechanisms in human detection behavior. Population responses in macaque V1 were recorded using voltage-sensitive dye imaging in fixating monkeys. Target detectability was analyzed from population responses at the retinotopic scale and at the scale of orientation columns using different decoding methods. At the retinotopic scale, a de-correlating filter (whitening filter) was used to derive spatial weighs to pool population responses (a doubly-whitened matched template model), and a simple matched-template model was used for estimating neural thresholds at the scale of orientation columns. To better estimate neural thresholds, the image data was corrected for motion artifacts. This correction increased the total number of imaging sessions that provided useful data.

Overall, I found that retinotopic thresholds rose at a faster rate than behavioral detection thresholds with increasing masking contrast (Figure 20). On the other hand, the rate of increase in neural thresholds from orientation column thresholds was less than behavioral thresholds with increasing masking contrast. Results from the relative trend in retinotopic performances suggest that the information conveyed at the retinotopic scale might be insufficient to perform detection in natural backgrounds. These results were

derived from a pooling rule that provides optimal performance for detecting a target on a uniform background, given the limits imposed by measurement noise and neural noise. A previous study of detection in uniform backgrounds found that optimal pooling at the retinotopic scale exceeded behavioral performances (Chen et al., 2006). However, the results from this thesis suggest that additional sources of neural information might be required to explain the monkeys' detection performance in natural image backgrounds.

The visual system might not be using an optimal pooling rule, and the relative performance in combining retinotopic signals across downstream areas could fall short of the optimal pooling rule, suggesting that detection sensitivities from retinotopic signals could decrease at a higher rate than behavioral detection sensitivities. The results of my psychophysical experiments are generally consistent with human contrast discrimination experiments, where researchers measured in humans the contrast increment to a pedestal contrast needed to produce a detectable difference (Foley & Legge, 1981; Nachmias & Sansbury, 1974). These psychophysical findings indicate that a progressively larger contrast increment is required for the target pattern in higher background contrasts. The neural basis of threshold elevation is unknown, but could be related to contrast nonlinearities that have been observed in the early visual system. The effects of these nonlinearities are evident in the contrast response function of V1 cells, where contrast discriminability is attenuated with higher stimulus contrast. The discriminability of contrast increments is decreased because of the compressive contrast response relationship in higher contrasts—stimulus evoked responses reside closer to the saturated region of the contrast response function, as shown in Figure 23. Higher background contrast responses (black dots) are distributed near the saturated region of the overall contrast response function, which is consistent with lower contrast detectability.

On the other hand, I find that orientation columns responses were less susceptible to background contrast masking in comparison to retinotopic responses. Orientation column responses were pooled with a matched-template model that selectively weighed orientation columns that preferred the orientation of the target pattern (horizontal Gabor). Columnar thresholds systematically increased with background contrast, but the rate of threshold elevation was less than retinotopic responses and behavioral performances.

A potential implication from this observation is that orientation columns reflect signals that are effective in rendering structural components of the retinal image across a range of masking contrasts. In effect, orientation columnar responses could provide structural information of a target pattern when performing detection in complex natural backgrounds. As a wide range of structural components exist in natural backgrounds, retinotopic responses of the target could be obscured from background components, and additional sources of information would be required to distinguish the target from background noise. Findings from this thesis indicate that orientation columns convey signals that could provide additional information of the target pattern, in addition to retinotopic signals that reflect the shape and location of the target pattern.

In summary, this thesis suggests that population activity at the retinotopic scale is one factor contributing to detection, but is insufficient to predict detection performances in natural backgrounds. Moreover, population responses at the scale of orientation columns are likely to provide additional sources of contextual information to identify a target pattern in natural backgrounds.

Furthermore, findings from this thesis provide physiological evidence that is relevant to predictions from the Retina-V1 model (Bradley et al, 2014). Model predictions suggest that human detection performances can be explained from well-known physiological properties of the retina and V1 and by pooling V1 responses in a near optimal

manner. In this thesis population responses were pooled at the output layers of the topographic map and orientation columns, and showed that neural performances systematically changed in a similar manner to behavioral detection performances in macaques. Although the exact pooling mechanism of V1 populations is still an open question, my findings suggest that the information conveyed in the scale of retinotopic map and orientation columns could be used to predict detection performances in natural image backgrounds.

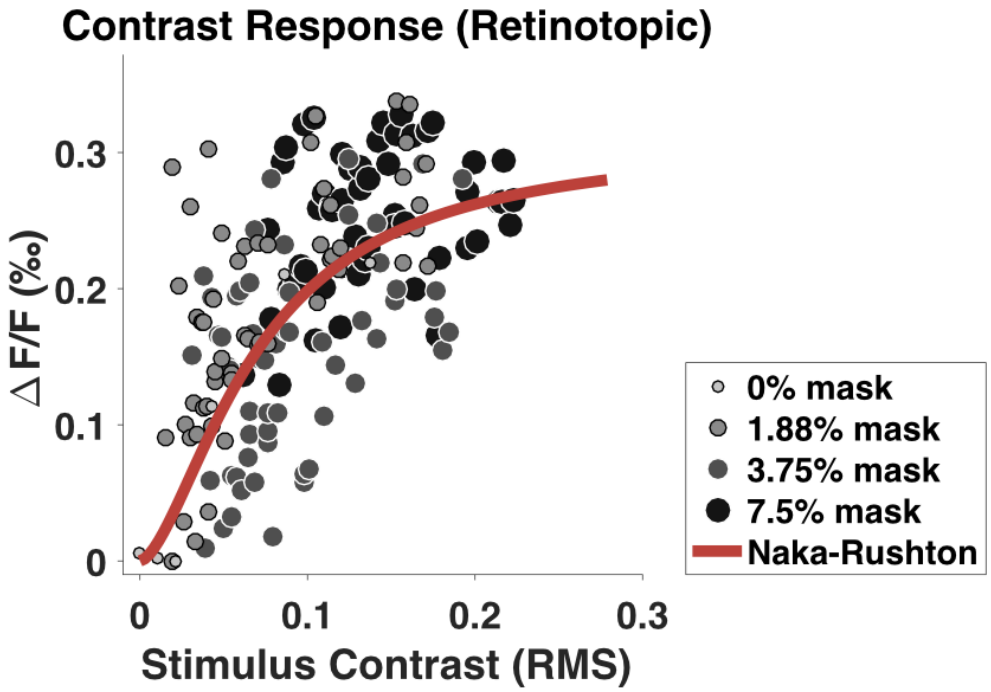

Figure 23: **Contrast response function of retinotopic responses.** VSDI responses from the center (peak) region of retinotopic responses are plotted as a function of stimulus contrast. Responses from different masking contrasts are color-coded, where darker dots represent higher background contrasts. Data are from experiments that were only included in previous analyses. Each data point is the average amplitude response from each condition within an imaging experiment. A single Naka-Rushton function is fitted to the data points in this plot using least-squares.

FUTURE DIRECTIONS

Fixating monkeys were used in this study as an exploratory attempt to link population responses to behavioral performances in detection. Ultimately, I seek to understand how populations work together to give rise to performing detection. To take a step closer to this question, physiological recordings would need to be conducted while the animal is performing the task in order to establish a direct link between neural mechanisms and behavior. Although this is beyond the scope of this thesis, an initial attempt was conducted in two imaging sessions using an additional monkey (monkey H). The monkey performed the exact same task from the psychophysical experiment (Chapter 2), and imaging data was acquired until the monkey broke fixation for reporting with a saccade (Figure 24A). Only two background contrast levels were tested at 3.75% and 7.5%. VSDI response profiles from the central area of the stimulus-evoked region are shown in Figure 24B. For each background contrast, the left plot depicts the mean amplitude and standard deviation, and the right plot shows the corresponding time course across target contrast. Overall, the quality of the data was too low to derive any thresholds. Imaging conditions were sub-optimal at the time of the experiments and could have compromised the data quality.

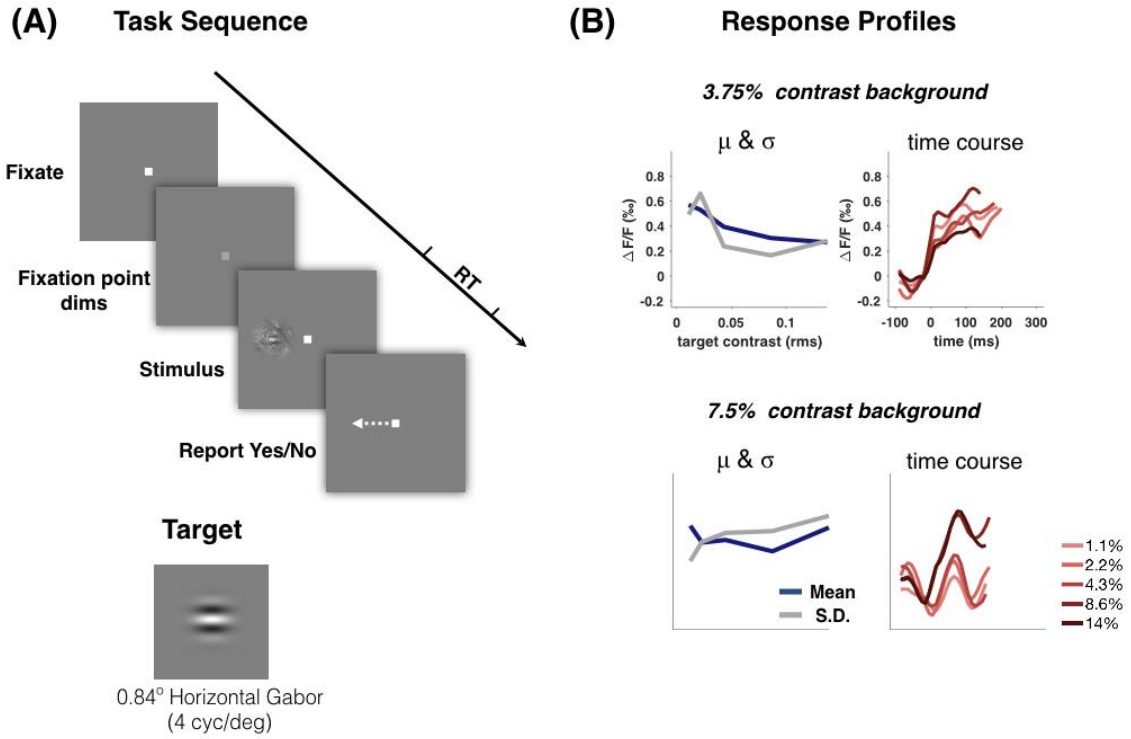

Figure 24: **Exploratory data from a detection task.** An additional monkey (monkey H) was used to perform a detection task in two imaging sessions. One session was for 3.75% contrast backgrounds and another for 7.5% contrast backgrounds. **(A)** Task sequence: same from the psychophysical experiment in Chapter 2—single interval reaction time task. **(B)** Response profiles from two imaging sessions. Top row: 3.75% backgrounds, bottom row, 3.75% backgrounds. The left subplot shows the mean and standard deviation of amplitudes across target contrasts. Response amplitudes are calculated from the central 1 x 1 mm^2 area of the stimulus-evoked region. The right subplot depicts the average time course before the monkey made a saccade for reporting. Time courses have different lengths according to the reaction time.

CONCLUDING REMARKS

In this thesis I studied how V1 population responses could be pooled to predict pattern behavioral detection performances in natural image backgrounds. I trained monkeys to perform a detection task using natural image backgrounds, and monitored the activity of populations of neurons in V1 from a separate pair of monkeys fixating to the

same stimuli. The results presented here suggest that the macaque is a good animal model in studying human detection behavior in natural scenes and detection performances could be predicted by pooling V1 populations based on first principles. As an exploratory attempt, populations were studied at the retinotopic and orientation columnar scales. At both scales, neural thresholds elevated with background contrast. Particularly, in natural image backgrounds, orientation columns appear to convey neural signals that are more effective in identifying a target pattern in comparison to retinotopic responses. A conjecture motivated by this finding is that orientation features contained in the orientation-column responses could be a reliable source of information in identifying a target pattern in a complex background. Furthermore, the visual system might be taking advantage of conveying orientation structures from the retinal image to downstream areas in orientation columns to perform a task. An important goal for a future work would be to monitor population responses while the monkey is performing the detection task. Linking neural decision variables to behavioral outcomes on a trial-by-trial basis is essential to further explore how V1 populations mediate behavioral outcomes in natural detection.

References

Arieli, A., Grinvald, A. & Slovin, H. Dural substitute for long-term imaging of cortical activity in behaving monkeys and its clinical implications. J. Neurosci. Methods 114,119–133 (2002).

Born, R. T., & Tootell, R. B. (1991). Single-unit and 2-deoxyglucose studies of side inhibition in macaque striate cortex. Proceedings of the National Academy of Sciences, 88(16), 7071-7075.

Bradley, C., Abrams, J., & Geisler, W. S. (2014). Retina - V1 Model of Detectability across the Visual Field The RV1 Model, 14, 1–28. http://doi.org/10.1167/14.12.22.doi

Cavanaugh, J. R., Bair, W., & Movshon, J. A. (2002). Nature and interaction of signals from the receptive field center and surround in macaque V1 neurons. Journal of neurophysiology, 88(5), 2530-2546.

Chen, Y., Geisler, W. S., & Seidemann, E. (2006). Optimal decoding of correlated neural population responses in the primate visual cortex. Nature Neuroscience, 9(11), 1412–20. http://doi.org/10.1038/nn1792

Chen, Y., Palmer, C. R., & Seidemann, E. (2012). The relationship between voltage-sensitive dye imaging signals and spiking activity of neural populations in primate V1. *Journal of neurophysiology*, *107*(12), 3281-3295.

Coen-Cagli, R., Kohn, A., & Schwartz, O. (2015). Flexible gating of contextual influences in natural vision. Nature Neuroscience, 18(11), 1648–1655. http://doi.org/10.1038/nn.4128

DeAngelis, G. C., Robson, J. G., Ohzawa, I., & Freeman, R. D. (1992). Organization of suppression in receptive fields of neurons in cat visual cortex. Journal of Neurophysiology, 68(1), 144-163.

DeValois, K.K., and Tootell, R.B.H. (1983). Spatial-frequency-spe- cific inhibition in cat striate cortex cells. J. Physiol. 336, 359–376.

Eckstein, M. P., Ahumada, A. J., & Watson, A. B. (1997). Visual signal detection in structured backgrounds. II. Effects of contrast gain control, background variations, and white noise. Journal of the Optical Society of America. A, Optics, Image Science, and Vision, 14(9), 2406–2419. http://doi.org/10.1364/JOSAA.14.002406

Felleman, Daniel J., and David C. Van Essen. "Distributed hierarchical processing in the primate cerebral cortex." Cerebral cortex 1.1 (1991): 1-47.

Finn, I. M., Priebe, N. J., & Ferster, D. (2007). The emergence of contrast- invariant orientation tuning in simple cells of cat visual cortex. Neuron, 54, 137–152.

Foley, J. M. (1994). Human luminance pattern-vision mechanisms: Masking experiments require a new model. Journal of the Optical Society of America A, 11(6), 1710-1719.

Geisler, W. (2008). Visual perception and the statistical properties of natural scenes. Annu. Rev. Psychol., 59, 167–192. http://doi.org/10.1146/annurev.psych.58.110405.085632

Geisler, W. S., & Diehl, R. L. (2003). A Bayesian approach to the evolution of perceptual and cognitive systems. Cognitive Science, 27(3), 379–402. http://doi.org/10.1016/S0364-0213(03)00009-0

Geisler, W.S., Perry, J.S., Super, B.J. & Gallogly, D.P. Edge co-occurrence in natural images predicts contour grouping performance. Vision Res. 41, 711–724 (2001)

Gilbert, C.D., and Wiesel, T.N. (1990). The influence of contextual stimuli on the orientation selectivity of cells in primary visual cortex of the cat. Vision Res. 30, 1689–1701.

Gilbert, C.D., and Wiesel, T.N. (1990). The influence of contextual stimuli on the orientation selectivity of cells in primary visual cortex of the cat. Vision Res. 30, 1689–1701.

Girshick, A. R., Landy, M. S., & Simoncelli, E. P. (2011). Cardinal rules: visual orientation perception reflects knowledge of environmental statistics. Nature Neuroscience, 14(7), 926–932. http://doi.org/10.1038/nn.2831

Goris, R. L. T., Putzeys, T., Wagemans, J., & Wichmann, F. a. (2013). A neural population model for visual pattern detection. Psychological Review, 120(3), 472–96. http://doi.org/10.1037/a0033136

Green DM, & Swets JA (1966) Signal Detection Theory and Psychophysics. New York: Wiley.

Grinvald A., Frostig R.D., Lieke E., Hildesheim R. (1988) Optical imaging of neuronal activity. Physiological reviews 68:1285-366.

Grinvald A., Hildesheim R. (2004) VSDI: a new era in functional imaging of cortical dynamics. Nature reviews. Neuroscience 5:874-85. DOI: 10.1038/nrn1536.

Horton, J. C., & Adams, D. L. (2005). The cortical column: a structure without a function. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *360*(1456), 837-862.

Hubel, D.H., and Wiesel, T.N. (1959). Receptive fields of single neurones in the cat's striate cortex. J. Physiol. 148, 574–591.

Knierim, J. J., & Van Essen, D. C. (1992). Neuronal responses to static texture patterns in area V1 of the alert macaque monkey. Journal of Neurophysiology, 67(4), 961-980.

Knill, D.C. & Richards, W. Perception as Bayesian Inference (Cambridge University Press, Cambridge, UK, 1996).

Michel, M. M., Chen, Y., Geisler, W. S., & Seidemann, E. (2013). An illusion predicted by V1 population activity implicates cortical topography in shape perception. Nature neuroscience, 16(10), 1477-1483.

Morrone, M. C., Burr, D. C., & Maffei, L. (1982). Functional implications of cross-orientation inhibition of cortical visual cells. I. Neurophysiological evidence. Proceedings of the Royal Society of London B: Biological Sciences, 216(1204), 335-354.

Nauhaus, I., Nielsen, K. J., Disney, A. A., & Callaway, E. M. (2012). Orthogonal micro-organization of orientation and spatial frequency in primate primary visual cortex. *Nature neuroscience*, *15*(12), 1683-1690.

Pelli, D. G. (1985). Uncertainty explains many aspects of visual contrast detection and discrimination. JOSA A, 2(9), 1508-1532.

Peterson, W. W. T. G., Birdsall, T., & Fox, W. (1954). The theory of signal detectability. *Transactions of the IRE professional group on information theory*, *4*(4), 171-212.

Polat, U., Mizobe, K., Pettet, M.W., Kasamatsu, T., and Norcia, A.M. (1998). Collinear stimuli regulate visual responses depending on cell's contrast threshold. Nature 391, 580–584.

Priebe, N. J., & Ferster, D. (2006). Mechanisms underlying cross-orientation suppression in cat visual cortex. Nature neuroscience, 9(4), 552-561.

Priebe, N. J., & Ferster, D. (2012). Mechanisms of Neuronal Computation in Mammalian Visual Cortex. Neuron, 75(2), 194–208. http://doi.org/10.1016/j.neuron.2012.06.011

Sato TK, Hausser M, Carandini M: Distal connectivity causes summation and division across mouse visual cortex. Nature Neuroscience 2014, 17:30-32.

Sceniak M, Ringach DL, Hawken M, Shapley R: Contrast's effect on spatial summation by macaque v1 neurons. Nat Neurosci 1999, 2:733-739.

Seidemann, E., Arieli, A., Grinvald, A. & Slovin, H. Dynamics of depolarization and hyperpolarization in the frontal cortex and saccade goal. Science 295, 862–865 (2002).

Shannon, C. E. (1949). Communication theory of secrecy systems. Bell system technical journal, 28(4), 656-715.

Shoham, D., Glaser, D. E., Arieli, A., Kenet, T., Wijnbergen, C., Toledo, Y., & Grinvald, A. (1999). Imaging cortical dynamics at high spatial and temporal resolution with novel blue voltage-sensitive dyes. Neuron, 24(4), 791-802.

Sit, Y. F., Chen, Y., Geisler, W. S., Miikkulainen, R., & Seidemann, E. (2009). Complex dynamics of V1 population responses explained by a simple gain-control model. Neuron, 64(6), 943-956.

Slovin, H., Arieli, A., Hildesheim, R. & Grinvald, A. Long-term voltage-sensitive dye imaging reveals cortical dynamics in behaving monkeys. J. Neurophysiol. 88, 3421–3438 (2002).

Tootell, R. B. H., Tsao, D., & Vanduffel, W. (2003). Neuroimaging weighs in: humans meet macaques in "primate" visual cortex. The Journal of Neuroscience: The Official Journal of the Society for Neuroscience, 23(10), 3981–3989. http://doi.org/23/10/3981

Walker, G.A., Ohzawa, I., and Freeman, R.D. (1999). Asymmetric suppression outside the classical receptive field of the visual cortex. J. Neurosci. 19, 10536–10553.

Watson, A. B., & Solomon, J. A. (1997). Model of visual contrast gain control and pattern masking. Journal of the Optical Society of America A, 14, 2379-2391.

Wickens, T. D. (2002). *Elementary signal detection theory*. Oxford University Press, USA.